EFFICIENCY-BASED HOUSING ALLOCATION: LEVERAGING DEA WITH CCR MODEL FOR ENHANCED DECISION – MAKING IN REAL ESTATES

Dissertation Submitted
in Partial Fulfilment of the Requirement for the Degree of

MASTERS OF SCIENCE

in

Mathematics

by

Ishika Gupta
(2K22/MSCMAT/58)

Himanshi
(2K22/MSCMAT/59)

Under the Supervision of
Prof. Anjana Gupta
Delhi Technological University

Department of Applied Mathematics

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi- 110042, India

May, 2024

ACKNOWLEDGEMENTS

The successful completion of any endeavor is always attributed to the individuals who have contributed, either implicitly or explicitly, to its progress. In the case of our project, the planning and research phases have reached fruition thanks to the dedicated efforts of numerous individuals. It is imperative to acknowledge their invaluable contributions in this report, as their involvement has been instrumental to our success. This expression of gratitude serves as a token of appreciation for their support in our undertaking. We extend our sincerest thanks to our project guide, whose guidance and solutions have consistently guided us through the complexities we encountered. Working under their leadership has been an honor and a pleasure. It is not only their knowledge that we have benefited from, but also their calm and composed demeanor, which has left an indelible impression on us. We remain indebted to them for these lessons, which will stay with us throughout our lives.

Furthermore, we would like to express our gratitude to Dr. ANJANA GUPTA (Professor) and RAMESH SRIVASTAVA (Professor), Head of the Department of Applied Mathematics, along with all the other faculty members who have imparted valuable knowledge and wisdom to us. We would like to acknowledge the focus of our work , which centers around the " Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate." The insights gained from this study have provided us with valuable insights into the effectiveness of these algorithms, contributing to the field of Data Envelopment Analysis.

Last but not least, we extend our heartfelt thanks to colleagues, friends, and parents, who have encouraged and inspired us throughout the project. Their unwavering support has been crucial in overcoming challenges and keeping us motivated. We are immensely grateful for their support and guidance, and we acknowledge their profound impact on our journey.

Thanking you

ISHIKA GUPTA AND HIMANSHI

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

We, **(Ishika Gupta, Himanshi), 2K22/MSCMAT/58, 2K22/MSCMAT/59**, hereby certify that the work which is presented in the dissertation entitled **Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate** in partial fulfillment of the requirements for the award of the Degree of M.Sc. Mathematics, submitted in the Department of **Applied Mathematics** , Delhi Technological University is an authentic record of my work carried out during the period from **August 2023** to **April 2024** under the supervision of **Dr. Anjana Gupta.**

The matter presented in the dissertation has not been submitted by us for the award of any other degree of this or any other institute.

Candidate's Signature

This is to certify that the students has incorporated all the corrections suggested by the examiners in the dissertation and the statement made by candidates is correct to the best of our knowledge.

Signature of Supervisor **Signature of External Examiner**

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Ishika Gupta, Himanshi, (2K22/MSCMAT/58, 2K22/MSCMAT/59)**, has carried out this research work presented in this dissertation entitled **Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate** for the award of the Degree of **M.Sc. Mathematics**, from the Department of Applied Mathematics , Delhi Technological University under my supervision. The dissertation embodied results of original work, and studies are carried out by the students themselves and the content of the dissertation do not form the basis for the award of any other degree to the candidates or to anybody else from this or any other University/ Institution.

Signature
Dr. Anjana Gupta
Professor

Date :

Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate

Ishika Gupta, Himanshi

ABSTRACT

This study introduces an innovative approach to optimizing house searches using Data Envelopment Analysis (DEA) with the Charnes, Kooper, and Rhodes (CCR) model. Utilizing live housing data from a diverse selection of properties, the research clusters houses based on price ranges and constructs efficiency frontier graphs to identify optimal housing allocations within each cluster. By quantifying the efficiency of houses, the DEA methodology provides a robust framework for nuanced comparisons and enhances decision-making in the real estate market.

The project investigates operational efficiency through the implementation of DEA, a powerful tool for comparing the efficiency of multiple units under varying conditions. DEA uses an efficiency frontier to signify peak performance achievable with specific inputs and outputs, offering insights into inefficiencies and opportunities for process optimization. Our methodology requires a comprehensive matrix of inputs, outputs, and relevant components for sampled decision-making units (DMUs), configured with specific metrics and orientation to provide relative efficiency scores and operational benchmarks.

Central to our analysis is the CCR model, which evaluates the efficiency of DMUs under the assumption of constant returns to scale, facilitating uniform comparisons and highlighting avenues for improvement. This approach aims to empower organizations by minimizing costs and maximizing benefits in various scenarios, such as goods transportation, service management, and process optimization. By considering a wide range of variables and potential conflicting goals, our study strives to enhance overall efficiency and inform decision-making, optimizing resource use and achieving high relative efficiency across diverse real-world contexts.

Keywords: Data Envelopment Analysis (DEA), Charnes Cooper and Rhodes (CCR), Efficiency Frontier Graph

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

DEA = Data Envelopment Analysis

CCR = Charnes Cooper Rhodes

DMU = Decision Making Units

EDA = Exploratory Data Analysis

x_{ij} = DMU Outputs

y_{ij} = DMU Inputs

u_r , v_i = variable weights

m = number of inputs

n = number of outputs

(i,j) = DMU arc

Chapter 1 : INTRODUCTION

Our project embarks on the investigation of operational efficiency and focuses on the implementation of Data Envelopment Analysis (DEA), a powerful methodology tailored for comparing the efficiency of multiple DMUs operating under various conditions. At its core, DEA presents conceptual units operating on the basis of an efficiency frontier, a dynamic scale that signifies the peak performance achievable with specific inputs and outputs. By examining units in relation to this boundary, DEA becomes a challenging tool that can identify inefficiencies and provide invaluable insights for improving processes and optimizing resource allocation.

1.1 Origin of Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a mathematical technique used to evaluate the relative efficiency of Decision-Making Units (DMUs) in various contexts, such as universities, banks, automobile manufacturers, educational institutions, hospitals, municipal corporations, and many more. It was initially introduced by Charnes, Cooper, and Rhodes in 1978, DEA has its roots in earlier work by Farrell and Fieldhouse in 1962. However, the linear programming model developed by Charnes, Cooper, and Rhodes (CCR) proved to be more flexible and effective compared to Farrell's method. The CCR model could be computed using standard linear programming software, offering computational efficiency[18]. This model created a vital link between a productivity index and efficiency measure, marking a significant advancement[18]. Notably, earlier contributions by agricultural economists in programming approaches for piecewise linear frontier production functions have largely been developed to model the efficient frontier in production. These methods focus on defining segments of linear functions that together form the frontier, representing the maximum possible output for a given set of inputs. overlooked until the CCR paper brought renewed attention and development to this area. Following this publication, more economists began utilizing this programming method, particularly for empirical studies, culminating in a comprehensive survey by Simar and Wilson in 2000.

The accuracy of DEA is influenced by the number of inputs, number of outputs and dimensions of the production set. To maintain a consistent level of estimation error, the sample size must increase exponentially with dimensionality. Due to this complexity, the only feasible approach to making inferences with DEA estimators involves using computationally intensive bootstrap methods, which require modifications with smoothing procedures to yield accurate results[18].

Conducting a DEA analysis requires a matrix containing inputs, outputs, and other pertinent data from the sample DMUs, after constructing the DEA model with specific metrics and orientation, this matrix is employed to solve the model[18]. The analysis results include relative efficiency scores and operational benchmarks for each DMU. Each DMU is assigned an efficiency score, denoted as 'e', and benchmarks or target values are computed to transform inefficient DMUs ($e < 1$) into efficient ones[18]. DEA aids in identifying potential enhancements for improving operational performance and differentiates between efficient and inefficient entities, the method involves establishing the efficiency frontier for the set of DMUs based on the observed data matrix and the DEA model[18]. This efficiency frontier defines the production possibility set, with DMUs on this frontier forming the reference set, DEA projects each DMU onto the efficiency frontier and identifies the maximum potential improvements for their inputs and/or outputs[18].

1.2 Motivation

The impetus for this project arises from the growing complexity and competitiveness across various sectors, which necessitates enhanced efficiency in resource allocation and performance optimization. Traditional efficiency assessment methods often fail to adequately capture the multifaceted nature of operations in diverse entities. Data Envelopment Analysis (DEA) provides a sophisticated and comprehensive method for estimating the efficiency of decision-making units (DMUs), considering multiple inputs and outputs without presuming a predefined functional relationship.

The primary objective of this project is to equip organizations and individuals with data-driven insights that enable informed decision-making and promote a culture of continuous improvement. The aim is to achieve optimal efficiency, balance costs and benefits, and maximize the utilization of available resources, thereby contributing to more effective and sustainable operations across various industries.

1.3 Dissertation Structure

To delve into the DEA analysis, our methodology requires a matrix including inputs, outputs and other components relevant to the sampled decision making units (DMUs), a DEA model, complexly configured with specific metrics and orientation, using relative efficiency scores and operational benchmarks. tailored to each DMU and offering a nuanced understanding of their performance.

Central to our endeavor is the Constant Returns to Scale (CCR) model within DEA, which primarily aims to evaluate the efficiency of DMUs, including entities such as firms, organizations, or any entity involved in the transformation of inputs into outputs. The CCR model, which operates under the assumption of a constant scale of operations, assumes that a doubling of inputs would lead to an equivalent doubling of outputs while maintaining a standardized scale. This basic assumption provides a uniform basis for evaluating effectiveness across entities, facilitating meaningful comparisons and illuminating avenues for operational improvement. Rooted in these methodologies, our project seeks to empower organizations in their quest to increase operational efficiency and make informed decisions regarding the use of resources.

The primary goal is to minimize costs and maximize benefits in a variety of real-world scenarios where resources must be allocated efficiently. Whether it is transporting goods, managing services or optimizing processes, the goal remains consistent – to strike a balance between minimizing costs and meeting specific requirements while ensuring maximum use of available capacities.

When solving a complex problem, traditional approaches can only consider the direct costs or benefits associated with specific tasks or operations. However, in a wider range of real-world scenarios, a multitude of variables can come into play for any decision-making process. In addition, decision makers may face conflicting goals where optimizing one aspect may compromise another. The ultimate goal is to increase overall efficiency, make informed decisions that take into account multiple objectives and variables, thereby optimizing the use of resources and achieving the highest possible relative efficiency in a variety of real-world contexts.

Chapter 2 : LITERATURE REVIEW

Year	Literature Review
An Analysis of Production as an Efficient Combination of Activities" Tjalling C. Koopmans.	According to Koopmans, an input-output vector is deemed efficiency is achieved when it is impossible to increase any output or decrease any input without simultaneously increasing other inputs or reducing other outputs. Koopmans described a feasible input-output vector as efficient based on this criterion. He demonstrated that a vector is considered efficient if only if it has a positive normal relative to the set of production possibilities[1].
(Debreu, 1951)	Koopmans provided a definition and description of efficiency, while Debreu's "coefficient of resource utilization" offered a way to measure it. Debreu quantified inefficiency by calculating a coefficient, that is derived by subtraction from one the maximum proportional reduction in all inputs needed to sustain the current level of output[2].
Farrell, 1957	Farrell acknowledged the influence of Koopmans and Debreu in his work. He laid the groundwork for new approaches to micro-level productivity and efficiency studies, offering fresh insights into the significance of productivity, efficiency, the calculation of standard technology and efficiency measurements[18]. Michael James Farrell's study, "The Measurement of Productive Efficiency," was a major inspiration for the CCR model and the preceding working papers.
C. Dennis Aigner Clairice D. Chu, 1968	C.Dennis Aigner and Clairice D. Chu's 1968 paper, "On Estimating the Industry Production Function," introduced a novel method for determining the production of industries. They proposed a deterministic and parametric framework using econometric techniques to estimate a production function, diverging from the previously sole average function method[18]. While recognizing Farrell's non-parametric approach, they chose the more traditional parametric method preferred by economists. They contended that Farrell's method was inadequate as it failed to accommodate various forms of production, including those adhering to the Law of Variable Proportions[18].

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985)	This book presents a thorough examination of the measurement of production efficiency. It discusses various efficiency measurement models, including parametric and non-parametric approaches, with a focus on DEA. The authors provide theoretical insights, mathematical formulations, and practical applications, making it an essential reference for understanding and implementing efficiency analysis in production.
Hsu, S. H., & Hwang, H. (2013)	This paper explores the application of DEA in environments with fuzzy production data, addressing non-efficiency scenarios. It introduces models that incorporate fuzzy logic to handle imprecise inputs and outputs, enhancing the robustness of efficiency measurements. The study includes case studies demonstrating the practical utility of these models in real-world production settings.
Kao, C., & Liu, S. T. (2000)	This study applies DEA to assess the efficiency of university libraries in Taiwan, dealing with missing data challenges. The authors propose methods to handle incomplete data without compromising the integrity of efficiency scores. Their approach ensures accurate and reliable performance evaluations, providing valuable insights into resource utilization and operational efficiency in educational institutions.
Cooper, W. William, Seiford, L. M., & Tone, K. (2007)	This detailed text provides an extensive overview of DEA, including models, applications, and software tools. It covers the evolution of DEA methodologies, theoretical underpinnings, and practical applications in multiple sectors. The book includes case studies, illustrative examples, and a DEA-solver software, serving as a valuable resource for advanced research and practical implementation.
Zhu, J. (2003)	This book offers a detailed introduction to performance evaluation and benchmarking using DEA, accompanied by spreadsheet-based tools. It covers theoretical foundations, model formulations, and practical examples across various industries. The integration of DEA with spreadsheets facilitates accessible and efficient performance analysis, making it a practical guide for both researchers and practitioners.

Cook, W. D., & Seiford, L. M. (2009)	This article reviews the advancements in DEA over thirty years, highlighting significant theoretical and methodological developments. It discusses the evolution of DEA models, their applications, and emerging trends in efficiency analysis. The paper provides a comprehensive summary of DEA's impact on operations research and management science, offering insights into future research directions.
Hollingsworth, B. (1999)	This paper discusses the application of DEA in evaluating public sector performance. It reviews methodologies, challenges, and case studies, illustrating how DEA can improve resource allocation and service delivery in public organizations. The study emphasizes the importance of efficiency measurement in public sector reform and policy-making, providing practical insights and recommendations.
Seiford, L. M., & Thrall, R. M. (1990)	This paper examines recent advancements in Data Envelopment Analysis (DEA), particularly in mathematical programming approaches to frontier analysis. It covers the latest developments in DEA models, methods for measuring efficiency, and their applications across different fields. The review emphasizes methodological innovations and the practical significance of DEA, offering a thorough overview of the current state-of-the-art in efficiency analysis.

Chapter 3 : PRELIMINARIES

3.1 Data Envelopment Analysis (DEA)

Data envelopment analysis (DEA) is a quantitative analysis technique which evaluates the efficiency of decision making units (DMUs) in transforming inputs into outputs [16]. Unlike traditional parametric methods, DEA do not require a predetermined functional relationships between the inputs and outputs, which makes it easy to analyze complex systems and evaluate performance without making specific assumptions about the production process[16].

The DEA was born in the late 1970s based on the research of Abraham Charnes, William W. Cooper, and Edward Rhodes. Their goal was to develop a method to measure the relative performance of organizations in situations where traditional approaches may not be applicable.

This method allows you to compare multiple units operating under different conditions by establishing a performance boundary that represents the efficient performance for the given inputs and outputs. By evaluating units in relation to this boundary, DEA identifies inefficient units and provides a framework for improving processes and resource allocation.

It is used in fields such as economics, operations research and management science. DEA is valuable precisely because it doesn't demand preconceived notions about the functional form of the production process. This attribute renders it particularly useful in scenarios where traditional econometric methods face limitations.

The primary aim of DEA is to gauge the relative efficiency of various Decision-Making Units (DMUs) in transforming multiple inputs into multiple outputs. These can be companies, organizations, departments or any entity that converts inputs into outputs. By comparing these units, DEA helps determine which units have the highest level of output relative to their input and serves as a benchmark for improving the performance of inefficient units.

To conduct a DEA analysis, one needs a matrix containing input, output, and complementary components of the DMU sample. Once the DEA model is constructed using specific criteria and orientation, matrices are utilized to solve the model, yielding relative efficiency scores and operational metrics for each DMU. Each DMU receives an efficiency score 'e', and target values, or benchmarks, are computed to transform inefficient DMUs ($e<1$) into efficient ones[18].

3.2 CCR Model

The CCR model in DEA stands for Charness, Cooper and Rhodes model, named after the three researchers who introduced it. DEA is a non-parametric mathematical method employed to assess the relative efficiencies of decision-making units across diverse domains like economics, operations research, and management.

The CCR model is one of the foundational models in DEA and was initially given by Charnes, Koper and Rhodes in 1978. The abbreviation CCR represents the initials of the researcher's last name.

The primary goal of the CCR model is to assess the efficiency of DMUs, which can be firms, organizations or any entity that transforms inputs into outputs. A DEA model requires inputs and output data for each DMU under consideration. Inputs represent the resources or the material used for production whereas outputs denote the products generated by the DMU[17].

The CCR model assumes that the volume of operations remains constant. This means that if a DMU doubles its inputs, it will also double its outputs while maintaining a

$$\text{Max } h_0 \quad \frac{\sum_{r=1}^{s} u_r}{\sum_{i=}^{m}}$$

Subject

$$\frac{\sum_{r=1}^{s}}{\sum_{i=} v_i x_i} \leq 1 \qquad j = 1, \dots\dots.,$$

$$v_r, u_i \geq 0 \qquad r = 1, \dots\dots\dots \qquad i = 1, \dots\dots\dots,$$

constant scale ratio. The CCR model provides an efficiency score for each DMU. An efficiency score of 1 indicates that the DMU is operating at maximum efficiency, making the best use of its inputs to produce output. A score below 1 indicates inefficiency and a score closer to 0, the DMU is considered less efficient.

When assessing the efficiency of a Decision-Making Unit (DMU), the outputs x_{ij} and inputs y_{rj} of the DMU are utilized, and optimization process is used to determine the variable weights u_r and v_i[18]. The reference set encompasses data on all production units, and their efficiency is evaluated relative to others. Efficiency ratings are incorporated in the optimization function and constraint, with the chosen production unit identified by the subscript '0' in the function while retaining its original subscripts in the constraint[18]. The optimization process aims to maximize the performance of the selected DMU by alloting it the most favorable weights permitted

$$\text{Max } h_0 = \sum_{r=1}^{s} u_r y_{r0}$$

Subject to:

$$\sum_{i=1}^{m} v_i x_{i0} = 1$$

$$\sum_{r=1}^{s} u_r y_{ij} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \qquad \text{j= 1,2,......n}$$

by the constraints[18].

In summary , the CCR Model in DEA provides a systematic approach to assess and compare the efficiency of decision making units, offering valuable insights for managerial decision making and resource allocation.

Chapter 4 : PROPOSED PAPER

We have a real time data of houses with various attributes. The dataset comprises **96,267** entries with various property-related attributes. The first column, "**House_code**", contains the code allotted to each house and the last column, "**category_code_description**," contains categorical information about the properties, presumably describing their types or classifications. The dataset includes numerical features such as **"exterior_condition," "frontage," "garage_spaces," "house_extension," "interior_condition," "market_value," "number_of_bathrooms," "number_of_bedrooms," "number_of_rooms," "number_stories," "sale_price," "taxable_building," "taxable_land," "total_area," "total_liveable_area," and "year_built."**

These features provide comprehensive details about each property, encompassing ph ysical characteristics like exterior and interior conditions, frontage, garage spaces, ho use extension, and the number of rooms and stories. Additionally, the dataset include s financial aspects such as market value, sale price, taxable building and land values. The "total_area" and "total_liveable_area" features offer insights into the overall and liveable spaces of the properties. Notably, the "year_built" column indicates the cons truction year of each property.

The objective is to assist buyers in maximizing features within a fixed budget. By pri oritizing criteria such as total liveable area, the number of rooms, bathrooms, frontag e, and other relevant factors, we aim to provide tailored recommendations. Our appro ach involves aligning the buyer's budget with the properties that offer the most exten sive set of desired features, ensuring a strategic match between financial constraints a nd aspirational housing attributes. Through this data-driven strategy, we seek to strea mline the decision-making process for buyers and enhance their overall satisfaction with their housing investment.

We start by initiating the data cleansing process, meticulously eliminating null values , removing outliers and filtering out extraneous information. Subsequently, we proce ed to partition the dataset into discrete clusters, driven by the market values of houses .

4.1 Exploratory Data Analysis (EDA) with Python

Step 1: Understanding the data

```
In [14]: print(df.head())
```

```
  category_code_description  exterior_condition  frontage  garage_spaces  \
0               Vacant Land                 0.0       0.0            0.0
1               Vacant Land                 NaN       0.0            0.0
2               Vacant Land                 NaN       0.0            0.0
3               Vacant Land                 NaN       0.0            0.0
4             Single Family                 1.0       0.0            2.0

   house_extension  interior_condition  market_value  number_of_bathrooms  \
0               30                 0.0             0                  0.0
1                0                 NaN         88800                  0.0
2                0                 NaN         88600                  0.0
3                0                 NaN         91000                  0.0
4                0                 1.0        537500                  3.0

   number_of_bedrooms  number_of_rooms  number_stories  sale_price  \
0                 0.0              0.0             0.0           1
1                 0.0              0.0             0.0           1
2                 0.0              0.0             0.0           1
3                 0.0              0.0             0.0           1
4                 3.0              9.0             3.0           1

   taxable_building  taxable_land  total_area  total_livable_area year_built
0                 0             0         0.0                 0.0        0.0
1                 0         88800         0.0                 0.0        0.0
2                 0         88600         0.0                 0.0        0.0
3                 0         91000         0.0                 0.0        0.0
4                 0        134400         0.0              2134.0     2019.0
```

Step 2 : Checking for null values in the data set

```
In [7]: df.isnull().sum()

Out[7]: category_code_description        0
        exterior_condition           25265
        frontage                         2
        garage_spaces                   21
        house_extension                  0
        interior_condition           26181
        market_value                     0
        number_of_bathrooms             10
        number_of_bedrooms              10
        number_of_rooms                 10
        number_stories                  10
        sale_price                       0
        taxable_building                 0
        taxable_land                     0
        total_area                       2
        total_livable_area              10
        year_built                      12
        dtype: int64
```

Graphical visualization of null values for every attribute.

```
In [12]: sns.heatmap(df.isnull(),yticklabels=False,cbar=False)

Out[12]: <Axes: >
```

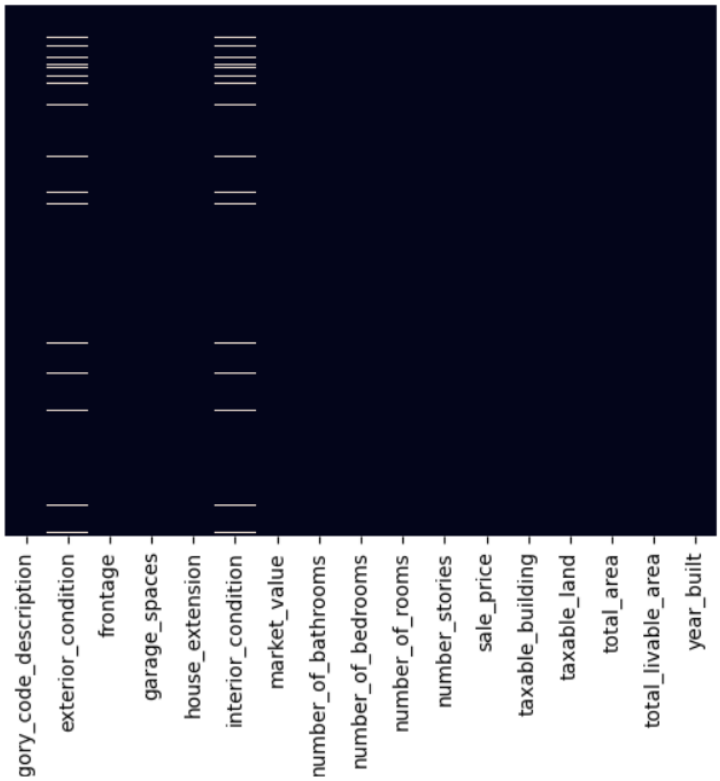


Figure 3.1 - Heat Map Depicting Null Values

<u>Step 3</u> : Treating all the null values

```
In [16]: df = df.dropna()
```

```
In [17]: df.isnull().sum()
```

```
Out[17]: category_code_description    0
         exterior_condition           0
         frontage                     0
         garage_spaces                0
         house_extension              0
         interior_condition           0
         market_value                 0
         number_of_bathrooms          0
         number_of_bedrooms           0
         number_of_rooms              0
         number_stories               0
         sale_price                   0
         taxable_building             0
         taxable_land                 0
         total_area                   0
         total_livable_area           0
         year_built                   0
         dtype: int64
```

<u>Step 4</u> : Code for removing outliers :

```
In [19]: df = df[df['total_area'] != 0]

In [20]: df = df[df['sale_price']!=0]

In [21]: df = df[df['market_value']!=0]

In [25]: df = df[df['frontage']!=0]

In [27]: df = df[df['number_of_rooms']!=0]

In [31]: df = df[df['number_of_bathrooms']!=0]

In [35]: df = df[df['total_livable_area'] != 0]

In [37]: df = df[df['year_built'] != 0]

In [39]: df = df[df['taxable_land'] != 0]
```

Step 5 : Summarizing the data :

In [49]:
```
max_sale = df['sale_price'].max()
min_sale = df['sale_price'].min()
print(max_sale)
print(min_sale)
```

```
30000000
10000
```

In [64]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 96267 entries, 34942 to 581323
Data columns (total 17 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   category_code_description  96267 non-null  object
 1   exterior_condition        96267 non-null  float64
 2   frontage                  96267 non-null  float64
 3   garage_spaces             96267 non-null  float64
 4   house_extension           96267 non-null  int64
 5   interior_condition        96267 non-null  float64
 6   market_value              96267 non-null  int64
 7   number_of_bathrooms       96267 non-null  float64
 8   number_of_bedrooms        96267 non-null  float64
 9   number_of_rooms           96267 non-null  float64
 10  number_stories            96267 non-null  float64
 11  sale_price                96267 non-null  int64
 12  taxable_building          96267 non-null  int64
 13  taxable_land              96267 non-null  int64
 14  total_area                96267 non-null  float64
 15  total_livable_area        96267 non-null  float64
 16  year_built                96267 non-null  object
dtypes: float64(10), int64(5), object(2)
memory usage: 13.2+ MB
```

The graphs visually depict the dataset and illustrate the relationships between various attributes.

The first scatter plot demonstrates how market value correlates with the number of rooms available within different categories identified by the **'Category_Code_Description'**.

We observe that single-family households tended to prefer houses with lower market values and fewer rooms, whereas multi-family households showed a preference for houses with slightly higher market values and a greater number of rooms.

```
In [33]: sns.scatterplot(data = df,x ='market_value',y = 'number_of_rooms',hue = 'category_code_description')
Out[33]: <Axes: xlabel='market_value', ylabel='number_of_rooms'>
```

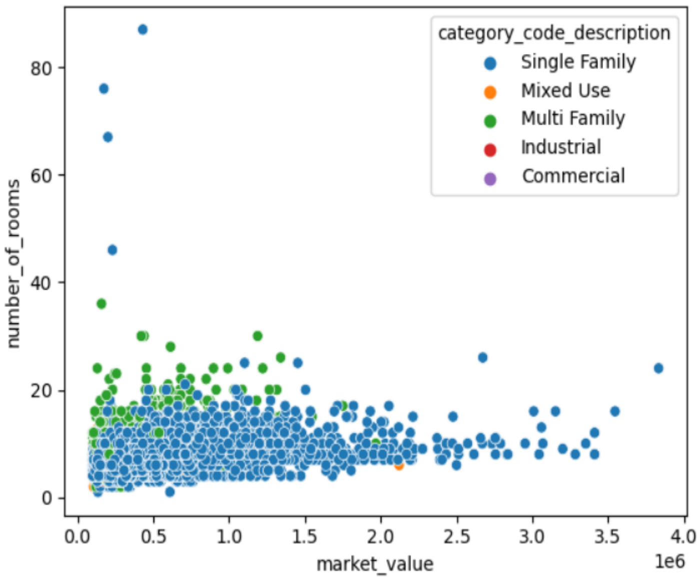


Figure 3.2 – Scatter Plot - Number of rooms versus Market values

Similarly, this scatter plot depicts the relation between market value and garage space. As garage space is one of the most important factor to be considered while buying the house.

```
In [37]: sns.scatterplot(data = df,x ='market_value',y = 'garage_spaces',hue='category_code_description' )
Out[37]: <Axes: xlabel='market_value', ylabel='garage_spaces'>
```

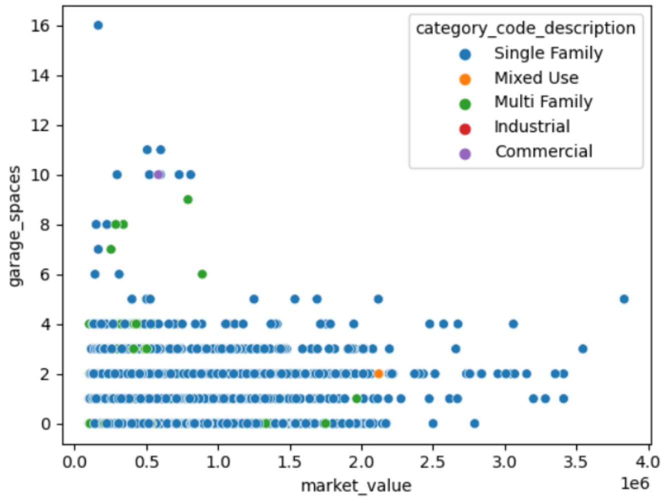


Figure 3.3 – Scatter Plot - Garage spaces versus Market values

This line plot depicts the interplay between market value and sales price. It highlights that as market value increases, sales prices become more variable, whereas lower market values correspond to lower sales prices, indicating a fluctuating trend.

In [36]:
```
sns.lineplot(data = df,x ='market_value',y = 'sale_price')
```
Out[36]:
```
<Axes: xlabel='market_value', ylabel='sale_price'>
```

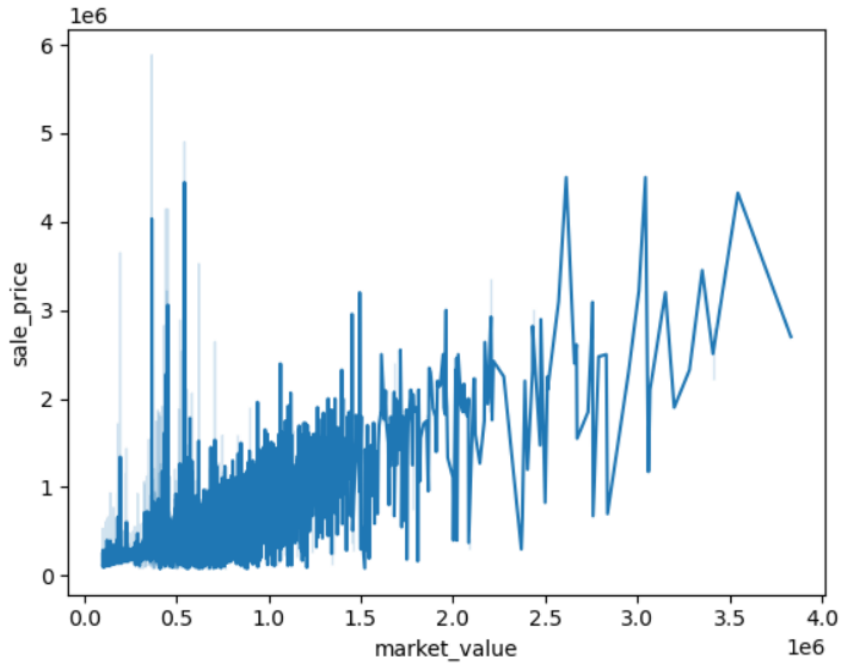


Figure 3.4- Line Plot - Sales price versus Market values

The following table describes the data through mean standard deviations, minimum values , maximum values and many more.

In [31]:
```
df.describe()
```
Out[31]:

	exterior_condition	frontage	garage_spaces	house_extension	interior_condition	market_value	number_of_bathrooms	number_of_bedrooms	n
count	96267.000000	96267.000000	96267.000000	96267.000000	96267.000000	9.626700e+04	96267.000000	96267.000000	
mean	3.541255	22.906792	0.517031	0.226235	3.523679	2.383480e+05	1.332024	3.177486	
std	0.853845	67.481269	0.593936	3.097478	0.860763	1.741180e+05	0.679122	0.807324	
min	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000e+05	1.000000	0.000000	
25%	3.000000	15.500000	0.000000	0.000000	3.000000	1.387000e+05	1.000000	3.000000	
50%	4.000000	16.330000	0.000000	0.000000	4.000000	1.895000e+05	1.000000	3.000000	
75%	4.000000	21.845000	1.000000	0.000000	4.000000	2.636000e+05	2.000000	3.000000	
max	7.000000	15917.000000	16.000000	99.000000	7.000000	3.832800e+06	21.000000	44.000000	

```
df.describe()
```

number_of_bathrooms	number_of_bedrooms	number_of_rooms	number_stories	sale_price	taxable_building	taxable_land	total_area	total_livable_area
96267.000000	96267.000000	96267.000000	96267.000000	9.626700e+04	9.626700e+04	9.626700e+04	9.626700e+04	9.626700e+04
1.332024	3.177486	6.352582	1.871171	2.222050e+05	1.404044e+05	5.296696e+04	2.279874e+03	1.486896e+03
0.679122	0.807324	1.319405	0.853525	2.631117e+05	1.105101e+05	5.166949e+04	6.969233e+03	7.900591e+03
1.000000	0.000000	1.000000	0.000000	8.000000e+04	0.000000e+00	6.900000e+01	1.000000e+00	2.590000e+02
1.000000	3.000000	6.000000	2.000000	1.200000e+05	8.654450e+04	2.399700e+04	9.920000e+02	1.136000e+03
1.000000	3.000000	6.000000	2.000000	1.650000e+05	1.150380e+05	3.651000e+04	1.504000e+03	1.320000e+03
2.000000	3.000000	7.000000	2.000000	2.470000e+05	1.654080e+05	6.213000e+04	2.207620e+03	1.600000e+03
21.000000	44.000000	87.000000	40.000000	3.000000e+07	2.682960e+06	1.149840e+06	1.591700e+06	2.445358e+06

4.2 Clustering Housing Data by Market Price for Budget-Based Segmentation

When searching for a new house, each person decides on a specific budget range for the house price. To accommodate this, we are dividing the data into different clusters based on market price ranges. This allows buyers to focus on houses within their budget. Within each price range, we can then suggest the best options for houses based on the efficiency calculated using Data Envelopment Analysis (DEA).

To determine the optimal number of clusters for our dataset based on market price and the number of houses within each price range, we are plotting a scatter graph between market price and the number of houses that fall within that price range. This graph provides valuable insights, such as highlighting that most houses fall under the price of $1,000,000, as well as other useful visual information.

```
In [21]: sorted_df = df.sort_values(by='market_value', ascending=True)
         import pandas as pd
         import matplotlib.pyplot as plt
         from sklearn.cluster import KMeans
         X = sorted_df[['market_value']]
         market_price_counts = sorted_df['market_value'].value_counts().sort_index()

         # Plot the scatter graph
         plt.figure(figsize=(10, 6))
         plt.scatter(market_price_counts.index, market_price_counts.values, c='blue', marker='o')
         plt.xlabel('Market Price')
         plt.ylabel('Number of Houses')
         plt.title('Scatter Plot: Market Price vs. Number of Houses')
         plt.grid(True)
         plt.show()
```

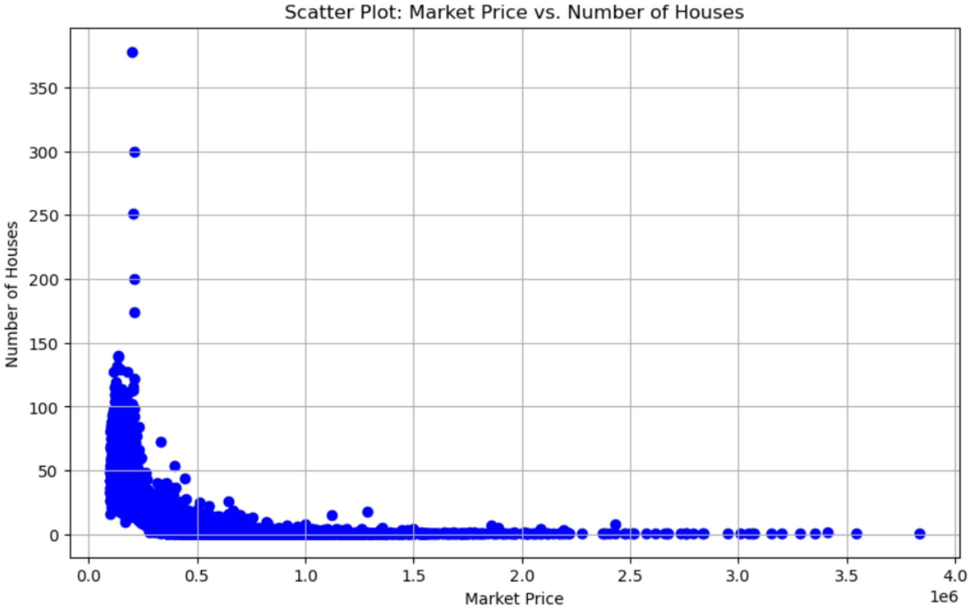


Figure 3.5 – Scatter Plot - Market price versus Number of Houses

Based on the scatter graph, we have determined that the optimal number of clusters for our dataset is six. Therefore, we are using K-Means clustering to create six clusters based on market price. The following code demonstrates this process by first fitting the K-Means algorithm to the data and then plotting a scatter graph to visualize the relationship between market price range and the number of houses in each cluster:

```
In [24]: optimal_clusters = 6
kmeans = KMeans(n_clusters=optimal_clusters, random_state=42)
sorted_df['cluster'] = kmeans.fit_predict(X)

# Now we can plot the scatter graph of market value and the number of houses
plt.figure(figsize=(10, 6))

# For the scatter plot, we will count the number of houses in each market price range (cluster)
cluster_counts = sorted_df['cluster'].value_counts().sort_index()
market_ranges = sorted_df.groupby('cluster')['market_value'].mean()

plt.scatter(market_ranges, cluster_counts, c='blue', marker='o')

plt.xlabel('Market Price Range (Average per Cluster)')
plt.ylabel('Number of Houses')
plt.title('Scatter Plot: Market Price Range vs. Number of Houses')
plt.grid(True)
plt.show()
```

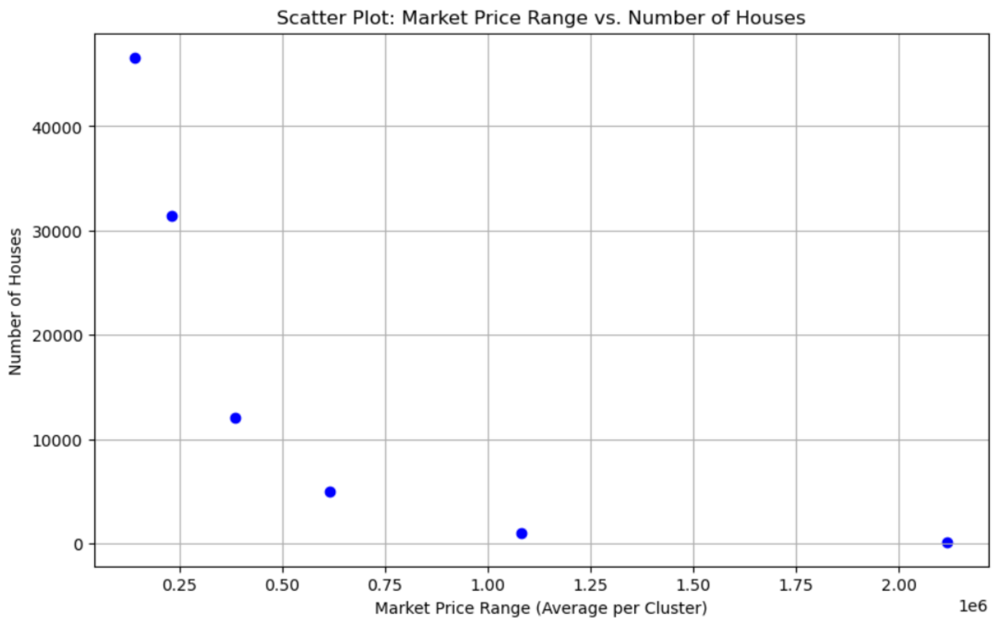


Figure 3.6 – Scatter Plot - Market price Range versus Number of houses

In this code, we first define the number of optimal clusters as six. We then apply the K-Means clustering algorithm to the dataset, creating six clusters based on market price. Each house is assigned a cluster label, which is added to the dataset as a new column. Following this, we prepare the data for the scatter plot by computing the count of houses in each cluster and the average market value for each cluster. Finally, we plot a scatter graph showing the average market price range against the number of houses in each cluster, providing a visual representation of the distribution of houses across different market price ranges.

The subsequent step involves segmenting the data into six distinct clusters based on the market value of houses. This segmentation facilitates the identification of budget-friendly market value clusters, simplifying the process for potential buyers without requiring extensive analysis of the entire dataset.

```
In [105]: group0 = (100000,150000 )
group1 = (150000,200000 )
group2 = (200000,300000 )
group3 = (300000, 600000)
group4 = (600000,1000000)
group5 = (1000000, 3832800)

# Separate data based on  groups
group0_data = df[(df['market_value'] >= group0[0]) & (df['market_value'] <= group0[1])]
group1_data = df[(df['market_value'] >= group1[0]) & (df['market_value'] <= group1[1])]
group2_data = df[(df['market_value'] >= group2[0]) & (df['market_value'] <= group2[1])]
group3_data = df[(df['market_value'] >= group3[0]) & (df['market_value'] <= group3[1])]
group4_data = df[(df['market_value'] >= group4[0]) & (df['market_value'] <= group4[1])]
group5_data = df[(df['market_value'] >= group5[0]) & (df['market_value'] <= group5[1])]
```

```
In [145]: output_file_path = 'Desktop\dtu msc sem1\(150000, 200000).xlsx'

          # Export the DataFrame to an Excel file
          group1_data.to_excel(output_file_path, index=False)
```

```
In [146]: output_file_path = 'Desktop\dtu msc sem1\(20000, 300000).xlsx'

          # Export the DataFrame to an Excel file
          group2_data.to_excel(output_file_path, index=False)
```

```
In [147]: output_file_path = 'Desktop\dtu msc sem1\(30000, g3).xlsx'

          # Export the DataFrame to an Excel file
          group3_data.to_excel(output_file_path, index=False)
```

```
In [148]: output_file_path = 'Desktop\dtu msc sem1\(150000, g40).xlsx'

          # Export the DataFrame to an Excel file
          group4_data.to_excel(output_file_path, index=False)
```

```
In [149]: output_file_path = 'Desktop\dtu msc sem1\(150000, g5).xlsx'

          # Export the DataFrame to an Excel file
          group5_data.to_excel(output_file_path, index=False)
```

4.3 DEA-Based Efficiency Analysis for Housing Clusters in R

Each cluster is addressed using the CCR model in R by employing the built-in library deaR. In this analysis, the decision-making units (DMUs) are denoted by their respective **`House_code`.** The attributes **`market_value`** and **`sales_price`** are designated as inputs, while the remaining attributes serve as outputs.

4.3.1 CLUSTER -1: (1,00,000 – 1,50,000)

Code -

```r
install.packages("deaR")
library(readxl)
data<-read_excel("C:\\Users\\HP\\Downloads\\(100000, g1).xlsx")
require(deaR)
ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
View(ccr_model)
result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
dmu_eval = 1:29751,dmu_ref = 1:29751)
efficiencies(result_ccr)
```

Output-

```
> install.packages("deaR")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and
install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/HP/AppData/Local/R/win-library/4.3’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/deaR_1.4.1.zip'
Content type 'application/zip' length 704537 bytes (688 KB)
downloaded 688 KB

package ‘deaR’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\HP\AppData\Local\Temp\RtmpmOMgVs\downloaded_packages
> library(readxl)
> data<-read_excel("C:\\Users\\HP\\Downloads\\(100000, g1).xlsx")
> require(deaR)
Loading required package: deaR
> ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
Warning messages:
1: 'read_data' is deprecated.
Use 'make_deadata' instead.
See help("Deprecated")
2: In make_deadata(datadea = datadea, ni = ni, no = no, dmus = dmus,  :
  There are negative or zero data. Try to translate the base point of the inputs/outputs with negati
ve data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
```

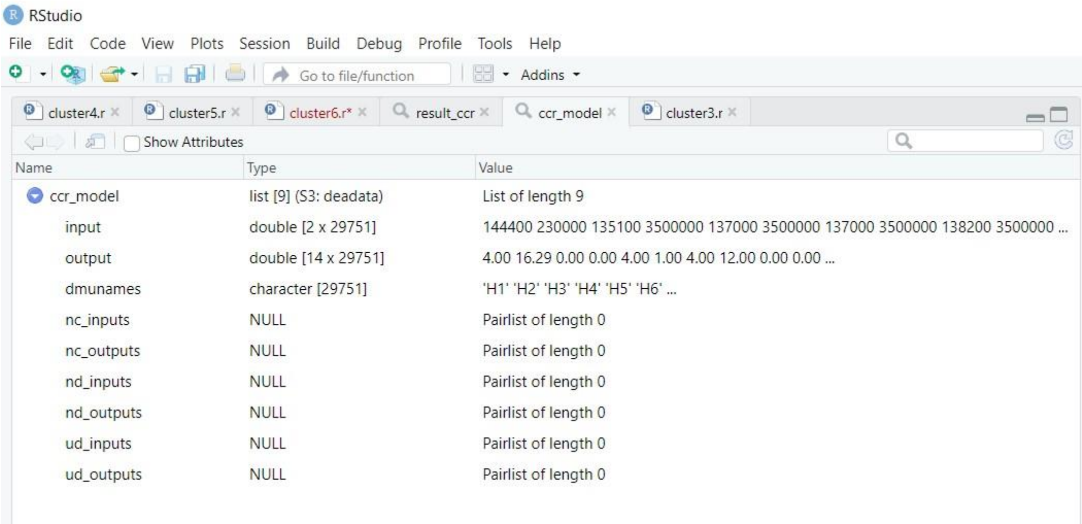
Console Terminal Background Jobs

R 4.3.2 · ~/

```
There are negative or zero data. Try to translate the base point of the inputs/outputs with negati
ve data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
> result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
+ dmu_eval = 1:740,dmu_ref = 1:740)
> efficiencies(result_ccr)
       H1        H2        H3        H4        H5        H6        H7        H8        H9 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H10       H11       H12       H13       H14       H15       H16       H17       H18 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.6990169 1.0000000 
      H19       H20       H21       H22       H23       H24       H25       H26       H27 
1.0000000 1.0000000 0.8661664 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H28       H29       H30       H31       H32       H33       H34       H35       H36 
1.0000000 1.0000000 0.9147651 0.8929352 0.9625297 1.0000000 1.0000000 1.0000000 1.0000000 
      H37       H38       H39       H40       H41       H42       H43       H44       H45 
0.8321980 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H46       H47       H48       H49       H50       H51       H52       H53       H54 
1.0000000 1.0000000 1.0000000 0.9467740 1.0000000 1.0000000 1.0000000 0.9625669 1.0000000 
      H55       H56       H57       H58       H59       H60       H61       H62       H63 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9529049 1.0000000 0.7349775 
      H64       H65       H66       H67       H68       H69       H70       H71       H72 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8223487 1.0000000 
      H73       H74       H75       H76       H77       H78       H79       H80       H81 
1.0000000 1.0000000 0.9810691 1.0000000 1.0000000 0.8930392 1.0000000 1.0000000 1.0000000 
      H82       H83       H84       H85       H86       H87       H88       H89       H90 
1.0000000 0.9190228 0.8765444 0.8790421 0.8649456 0.7488862 1.0000000 1.0000000 1.0000000 
      H91       H92       H93       H94       H95       H96       H97       H98       H99 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8158889 1.0000000 1.0000000 
     H100      H101      H102      H103      H104      H105      H106      H107      H108 
```

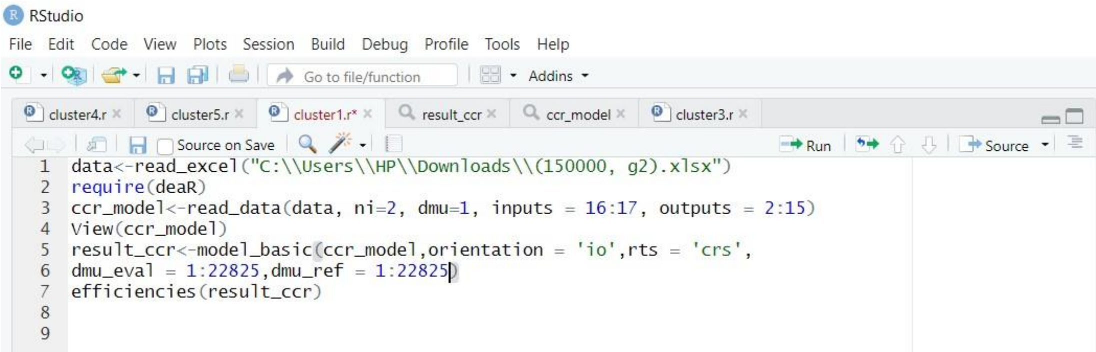
```
   H29627    H29628    H29629    H29630    H29631    H29632    H29633    H29634    H29635 
0.7155387 1.0000000 1.0000000 0.9124665 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
   H29636    H29637    H29638    H29639    H29640    H29641    H29642    H29643    H29644 
1.0000000 0.9080141 0.7835541 1.0000000 1.0000000 0.9494994 1.0000000 1.0000000 1.0000000 
   H29645    H29646    H29647    H29648    H29649    H29650    H29651    H29652    H29653 
0.7797121 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8074473 
   H29654    H29655    H29656    H29657    H29658    H29659    H29660    H29661    H29662 
0.8753508 0.8194343 1.0000000 1.0000000 0.8187682 1.0000000 1.0000000 1.0000000 1.0000000 
   H29663    H29664    H29665    H29666    H29667    H29668    H29669    H29670    H29671 
1.0000000 1.0000000 1.0000000 0.9973871 0.8558382 1.0000000 1.0000000 0.9668439 0.8351069 
   H29672    H29673    H29674    H29675    H29676    H29677    H29678    H29679    H29680 
1.0000000 1.0000000 1.0000000 1.0000000 0.9114207 1.0000000 1.0000000 1.0000000 0.7703737 
   H29681    H29682    H29683    H29684    H29685    H29686    H29687    H29688    H29689 
1.0000000 1.0000000 0.9070533 0.8100299 1.0000000 1.0000000 1.0000000 1.0000000 0.8385116 
   H29690    H29691    H29692    H29693    H29694    H29695    H29696    H29697    H29698 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9174681 1.0000000 1.0000000 1.0000000 
   H29699    H29700    H29701    H29702    H29703    H29704    H29705    H29706    H29707 
1.0000000 1.0000000 1.0000000 1.0000000 0.8912213 0.7758289 0.7437193 1.0000000 1.0000000 
   H29708    H29709    H29710    H29711    H29712    H29713    H29714    H29715    H29716 
1.0000000 1.0000000 1.0000000 1.0000000 0.9220733 1.0000000 1.0000000 1.0000000 1.0000000 
   H29717    H29718    H29719    H29720    H29721    H29722    H29723    H29724    H29725 
0.8006071 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8116811 1.0000000 0.8866227 
   H29726    H29727    H29728    H29729    H29730    H29731    H29732    H29733    H29734 
1.0000000 1.0000000 1.0000000 0.8999465 1.0000000 0.8434134 0.9297246 0.7600307 1.0000000 
   H29735    H29736    H29737    H29738    H29739    H29740    H29741    H29742    H29743 
0.8163053 1.0000000 0.8668065 0.8979791 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
   H29744    H29745    H29746    H29747    H29748    H29749    H29750    H29751 
1.0000000 1.0000000 0.9292144 0.8865323 1.0000000 1.0000000 1.0000000 1.0000000 
> 
```

View Model-



<u>4.3.2 CLUSTER -2: (1,50,000 – 2,00,000)</u>

<u>Code -</u>



Output-

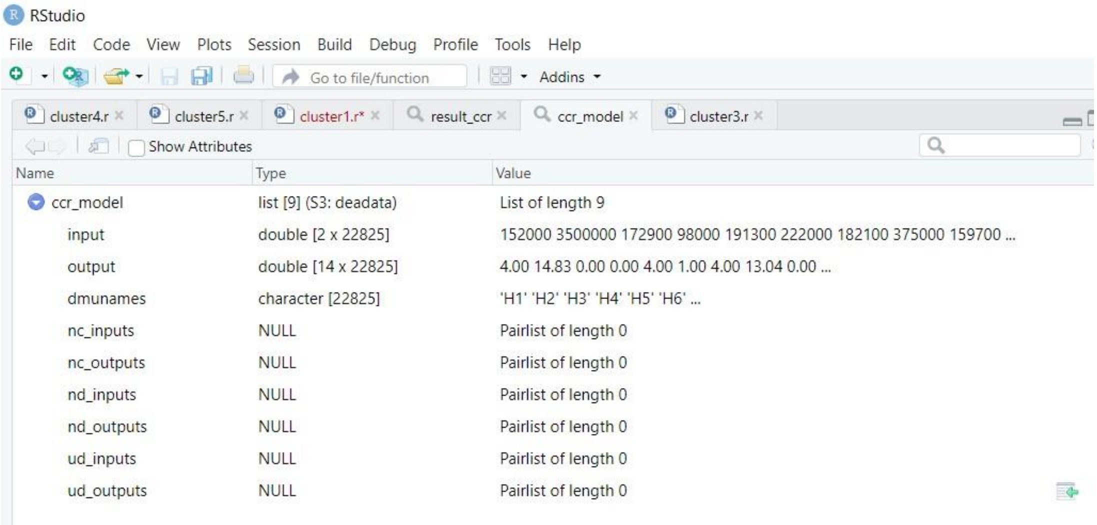
Console Terminal Background Jobs

R 4.3.2 · ~/

```
> data<-read_excel("C:\\Users\\HP\\Downloads\\(150000, g2).xlsx")
> require(deaR)
> ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
Warning messages:
1: 'read_data' is deprecated.
Use 'make_deadata' instead.
See help("Deprecated")
2: In make_deadata(datadea = datadea, ni = ni, no = no, dmus = dmus,  :
  There are negative or zero data. Try to translate the base point of the inputs/outputs with negati
ve data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
> result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
+ dmu_eval = 1:600,dmu_ref = 1:600)
> efficiencies(result_ccr)
       H1        H2        H3        H4        H5        H6        H7        H8        H9 
1.0000000 1.0000000 0.8712388 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H10       H11       H12       H13       H14       H15       H16       H17       H18 
1.0000000 0.9205307 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H19       H20       H21       H22       H23       H24       H25       H26       H27 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H28       H29       H30       H31       H32       H33       H34       H35       H36 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9112306 1.0000000 
      H37       H38       H39       H40       H41       H42       H43       H44       H45 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H46       H47       H48       H49       H50       H51       H52       H53       H54 
0.9216712 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
      H55       H56       H57       H58       H59       H60       H61       H62       H63 
1.0000000 1.0000000 0.9283546 0.8681756 0.8359475 1.0000000 0.9110998 1.0000000 1.0000000 
```

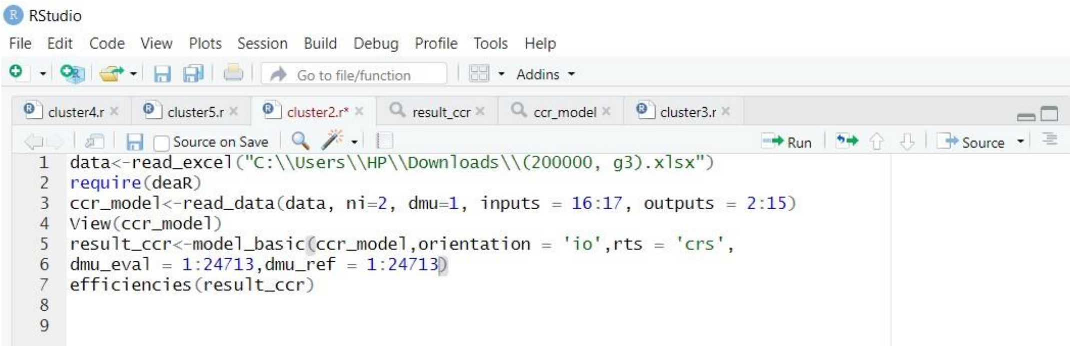
```
   H22700    H22701    H22702    H22703    H22704    H22705    H22706    H22707    H22708 
1.0000000 1.0000000 1.0000000 0.8884959 0.9781761 1.0000000 1.0000000 1.0000000 0.9738515 
   H22709    H22710    H22711    H22712    H22713    H22714    H22715    H22716    H22717 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9915170 1.0000000 0.7834608 
   H22718    H22719    H22720    H22721    H22722    H22723    H22724    H22725    H22726 
0.8670859 0.7894682 0.9094585 1.0000000 1.0000000 0.8218638 0.9347650 0.9743453 1.0000000 
   H22727    H22728    H22729    H22730    H22731    H22732    H22733    H22734    H22735 
1.0000000 0.9238971 1.0000000 1.0000000 1.0000000 0.9372171 1.0000000 0.8471395 1.0000000 
   H22736    H22737    H22738    H22739    H22740    H22741    H22742    H22743    H22744 
1.0000000 1.0000000 1.0000000 0.9526645 0.9521159 1.0000000 1.0000000 1.0000000 1.0000000 
   H22745    H22746    H22747    H22748    H22749    H22750    H22751    H22752    H22753 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9762793 1.0000000 1.0000000 1.0000000 
   H22754    H22755    H22756    H22757    H22758    H22759    H22760    H22761    H22762 
0.8446825 0.8458319 0.8676953 1.0000000 0.8202051 1.0000000 1.0000000 1.0000000 0.9097295 
   H22763    H22764    H22765    H22766    H22767    H22768    H22769    H22770    H22771 
1.0000000 1.0000000 0.9397899 0.9948754 1.0000000 1.0000000 0.9255926 1.0000000 0.9055166 
   H22772    H22773    H22774    H22775    H22776    H22777    H22778    H22779    H22780 
0.8209417 1.0000000 0.8082995 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9581779 
   H22781    H22782    H22783    H22784    H22785    H22786    H22787    H22788    H22789 
0.9909234 0.7780756 0.8580980 1.0000000 0.7903621 0.9628365 1.0000000 1.0000000 1.0000000 
   H22790    H22791    H22792    H22793    H22794    H22795    H22796    H22797    H22798 
1.0000000 1.0000000 1.0000000 0.9664412 0.9098084 1.0000000 1.0000000 0.9196503 1.0000000 
   H22799    H22800    H22801    H22802    H22803    H22804    H22805    H22806    H22807 
0.9469679 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
   H22808    H22809    H22810    H22811    H22812    H22813    H22814    H22815    H22816 
0.9849465 0.9182783 1.0000000 0.9275563 0.7985113 0.9969737 1.0000000 0.8433550 1.0000000 
   H22817    H22818    H22819    H22820    H22821    H22822    H22823    H22824    H22825 
1.0000000 0.8453933 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8477372 1.0000000 
> 
```

View Model-

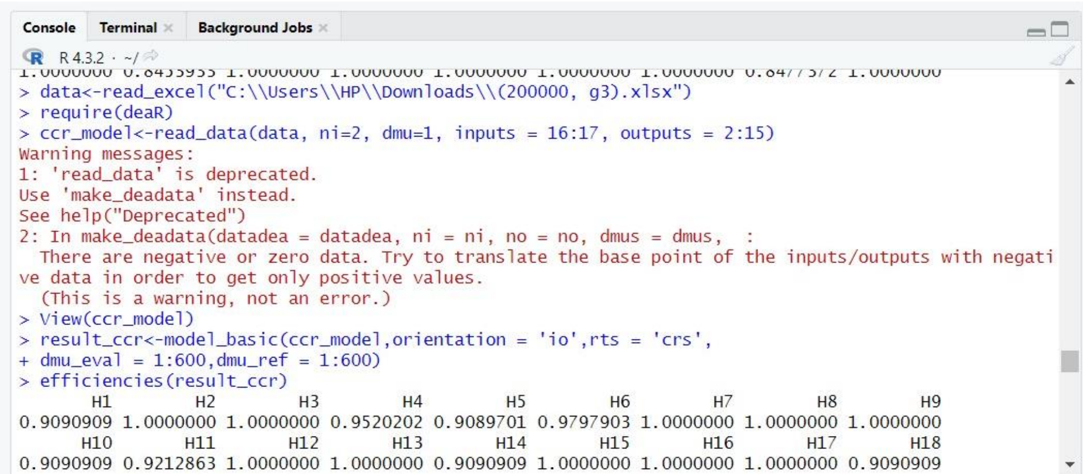


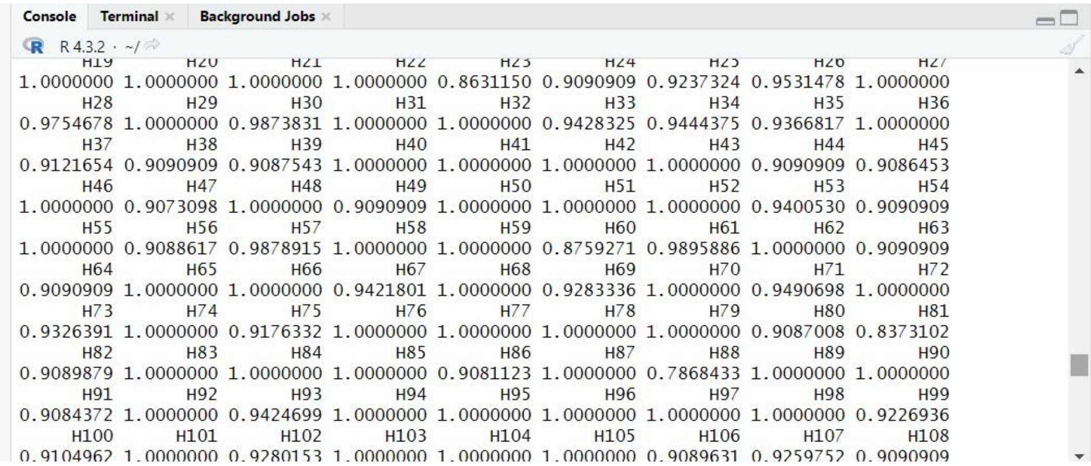
4.3.3 CLUSTER -3: (2,00,000 – 3,00,000)

Code -

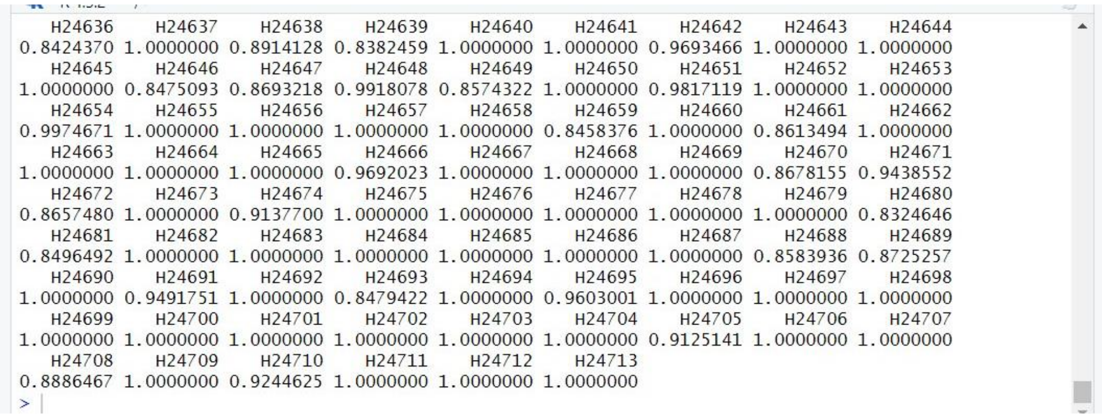


Output-



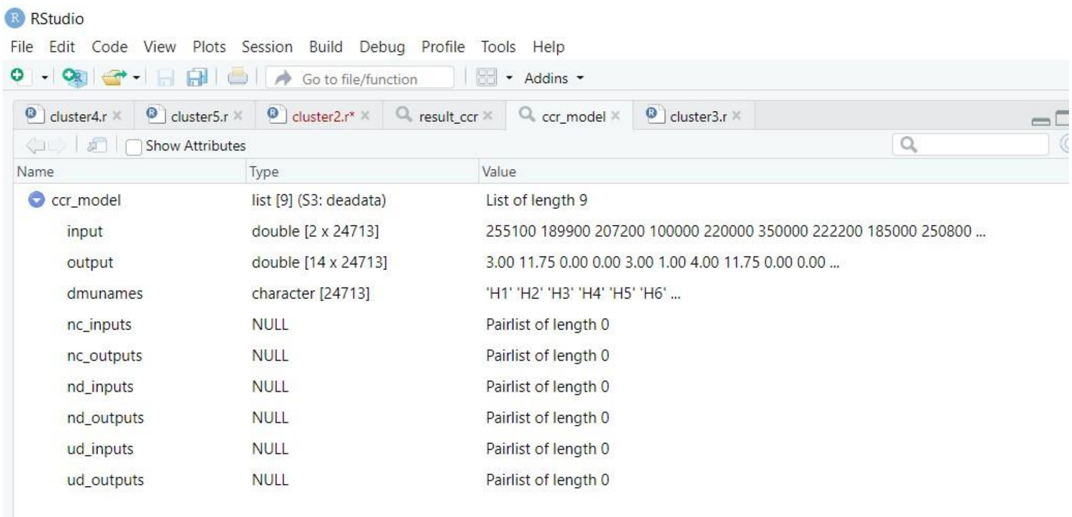


```
Console  Terminal  Background Jobs
R 4.3.2 · ~/
      H19       H20       H21       H22       H23       H24       H25       H26       H27
1.0000000 1.0000000 1.0000000 1.0000000 0.8631150 0.9090909 0.9237324 0.9531478 1.0000000
      H28       H29       H30       H31       H32       H33       H34       H35       H36
0.9754678 1.0000000 0.9873831 1.0000000 1.0000000 0.9428325 0.9444375 0.9366817 1.0000000
      H37       H38       H39       H40       H41       H42       H43       H44       H45
0.9121654 0.9090909 0.9087543 1.0000000 1.0000000 1.0000000 1.0000000 0.9090909 0.9086453
      H46       H47       H48       H49       H50       H51       H52       H53       H54
1.0000000 0.9073098 1.0000000 0.9090909 1.0000000 1.0000000 1.0000000 0.9400530 0.9090909
      H55       H56       H57       H58       H59       H60       H61       H62       H63
1.0000000 0.9088617 0.9878915 1.0000000 1.0000000 0.8759271 0.9895886 1.0000000 0.9090909
      H64       H65       H66       H67       H68       H69       H70       H71       H72
0.9090909 1.0000000 1.0000000 0.9421801 1.0000000 0.9283336 1.0000000 0.9490698 1.0000000
      H73       H74       H75       H76       H77       H78       H79       H80       H81
0.9326391 1.0000000 0.9176332 1.0000000 1.0000000 1.0000000 1.0000000 0.9087008 0.8373102
      H82       H83       H84       H85       H86       H87       H88       H89       H90
0.9089879 1.0000000 1.0000000 1.0000000 0.9081123 1.0000000 0.7868433 1.0000000 1.0000000
      H91       H92       H93       H94       H95       H96       H97       H98       H99
0.9084372 1.0000000 0.9424699 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9226936
     H100      H101      H102      H103      H104      H105      H106      H107      H108
0.9104962 1.0000000 0.9280153 1.0000000 1.0000000 1.0000000 0.9089631 0.9259752 0.9090909
```



```
   H24636    H24637    H24638    H24639    H24640    H24641    H24642    H24643    H24644
0.8424370 1.0000000 0.8914128 0.8382459 1.0000000 1.0000000 0.9693466 1.0000000 1.0000000
   H24645    H24646    H24647    H24648    H24649    H24650    H24651    H24652    H24653
1.0000000 0.8475093 0.8693218 0.9918078 0.8574322 1.0000000 0.9817119 1.0000000 1.0000000
   H24654    H24655    H24656    H24657    H24658    H24659    H24660    H24661    H24662
0.9974671 1.0000000 1.0000000 1.0000000 1.0000000 0.8458376 1.0000000 0.8613494 1.0000000
   H24663    H24664    H24665    H24666    H24667    H24668    H24669    H24670    H24671
1.0000000 1.0000000 1.0000000 0.9692023 1.0000000 1.0000000 1.0000000 0.8678155 0.9438552
   H24672    H24673    H24674    H24675    H24676    H24677    H24678    H24679    H24680
0.8657480 1.0000000 0.9137700 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8324646
   H24681    H24682    H24683    H24684    H24685    H24686    H24687    H24688    H24689
0.8496492 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.8583936 0.8725257
   H24690    H24691    H24692    H24693    H24694    H24695    H24696    H24697    H24698
1.0000000 0.9491751 1.0000000 0.8479422 1.0000000 0.9603001 1.0000000 1.0000000 1.0000000
   H24699    H24700    H24701    H24702    H24703    H24704    H24705    H24706    H24707
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9125141 1.0000000 1.0000000
   H24708    H24709    H24710    H24711    H24712    H24713
0.8886467 1.0000000 0.9244625 1.0000000 1.0000000 1.0000000
>
```

View Model-



Name	Type	Value
ccr_model	list [9] (S3: deadata)	List of length 9
input	double [2 x 24713]	255100 189900 207200 100000 220000 350000 222200 185000 250800 ...
output	double [14 x 24713]	3.00 11.75 0.00 0.00 3.00 1.00 4.00 11.75 0.00 0.00 ...
dmunames	character [24713]	'H1' 'H2' 'H3' 'H4' 'H5' 'H6' ...
nc_inputs	NULL	Pairlist of length 0
nc_outputs	NULL	Pairlist of length 0
nd_inputs	NULL	Pairlist of length 0
nd_outputs	NULL	Pairlist of length 0
ud_inputs	NULL	Pairlist of length 0
ud_outputs	NULL	Pairlist of length 0

4.3.4 CLUSTER -4: (3,00,000 – 6,00,000)

Code -

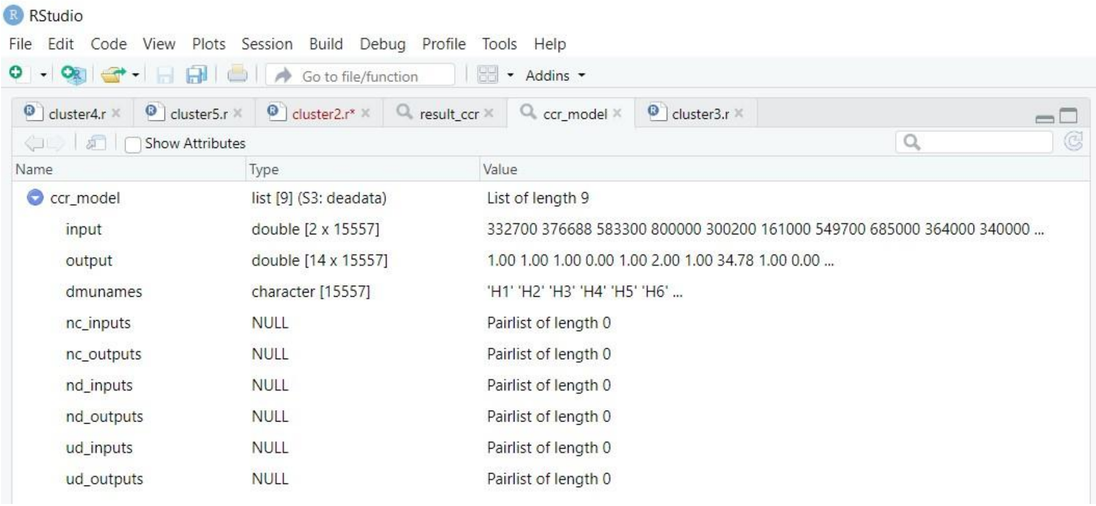
```
RStudio
File  Edit  Code  View  Plots  Session  Build  Debug  Profile  Tools  Help
Go to file/function   Addins
cluster4.r  cluster5.r  cluster4.r*  result_ccr  ccr_model  cluster3.r
Source on Save   Run   Source

data<-read_excel("C:\\Users\\HP\\Downloads\\(300000, g4).xlsx")
require(deaR)
ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
View(ccr_model)
result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
dmu_eval = 1:15557,dmu_ref = 1:15557)
efficiencies(result_ccr)
```

Output-

```
> data<-read_excel("C:\\Users\\HP\\Downloads\\(300000, g4).xlsx")
> require(deaR)
> ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
Warning messages:
1: 'read_data' is deprecated.
Use 'make_deadata' instead.
See help("Deprecated")
2: In make_deadata(datadea = datadea, ni = ni, no = no, dmus = dmus,  :
  There are negative or zero data. Try to translate the base point of the inputs/outputs with negati
ve data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
> result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
+ dmu_eval = 1:600,dmu_ref = 1:600)
> efficiencies(result_ccr)
       H1        H2        H3        H4        H5        H6        H7        H8        H9 
1.0000000 1.0000000 1.0000000 0.9256840 1.0000000 1.0000000 0.9951674 1.0000000 1.0000000 
      H10       H11       H12       H13       H14       H15       H16       H17       H18 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9107118 1.0000000 0.9683072 1.0000000 
      H19       H20       H21       H22       H23       H24       H25       H26       H27 
1.0000000 0.9115155 0.9529113 0.9147472 0.8510604 0.8510604 0.8396037 0.8510604 0.8510604 
      H28       H29       H30       H31       H32       H33       H34       H35       H36 
0.8510604 0.8510604 0.8510604 0.8510604 0.8510604 0.8595833 0.8595833 0.8595833 0.8595833 
      H37       H38       H39       H40       H41       H42       H43       H44       H45 
0.8595833 0.8595833 0.8595833 0.8595833 0.8595833 0.8595833 1.0000000 0.9904117 0.9548502 
      H46       H47       H48       H49       H50       H51       H52       H53       H54 
0.8684668 1.0000000 1.0000000 0.8918218 0.9353663 1.0000000 0.9617778 1.0000000 0.8760952 
      H55       H56       H57       H58       H59       H60       H61       H62       H63 
1.0000000 1.0000000 0.9219177 0.9835577 0.9800234 0.9053477 1.0000000 0.9661194 0.9194801 
      H64       H65       H66       H67       H68       H69       H70       H71       H72 
1.0000000 0.9619615 1.0000000 0.8962011 0.9181934 1.0000000 1.0000000 0.9794075 1.0000000 
```

```
   H15472    H15473    H15474    H15475    H15476    H15477    H15478    H15479    H15480 
1.0000000 0.9080319 0.9175371 1.0000000 1.0000000 0.8902707 1.0000000 1.0000000 1.0000000 
   H15481    H15482    H15483    H15484    H15485    H15486    H15487    H15488    H15489 
0.9249875 0.9977578 0.9700726 1.0000000 0.9174616 1.0000000 0.9638213 0.9102838 1.0000000 
   H15490    H15491    H15492    H15493    H15494    H15495    H15496    H15497    H15498 
1.0000000 0.9683181 1.0000000 0.8828125 1.0000000 0.9462343 1.0000000 0.9202976 1.0000000 
   H15499    H15500    H15501    H15502    H15503    H15504    H15505    H15506    H15507 
1.0000000 1.0000000 1.0000000 0.9716294 0.9066390 0.9167592 1.0000000 1.0000000 1.0000000 
   H15508    H15509    H15510    H15511    H15512    H15513    H15514    H15515    H15516 
1.0000000 1.0000000 0.9011206 1.0000000 0.9136773 0.9433320 1.0000000 1.0000000 0.9146110 
   H15517    H15518    H15519    H15520    H15521    H15522    H15523    H15524    H15525 
0.9244966 1.0000000 1.0000000 1.0000000 1.0000000 0.9470944 1.0000000 1.0000000 0.9773487 
   H15526    H15527    H15528    H15529    H15530    H15531    H15532    H15533    H15534 
1.0000000 0.9296145 0.9099425 1.0000000 1.0000000 1.0000000 0.9548379 0.8812352 1.0000000 
   H15535    H15536    H15537    H15538    H15539    H15540    H15541    H15542    H15543 
0.9481809 0.8932638 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
   H15544    H15545    H15546    H15547    H15548    H15549    H15550    H15551    H15552 
1.0000000 1.0000000 0.9777504 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
   H15553    H15554    H15555    H15556    H15557 
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
> 
```

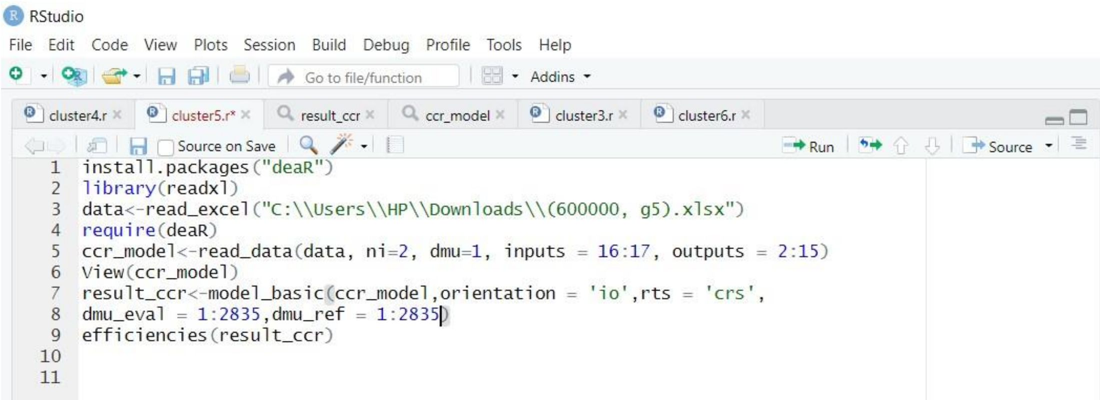
View Model-



Name	Type	Value
ccr_model	list [9] (S3: deadata)	List of length 9
input	double [2 x 15557]	332700 376688 583300 800000 300200 161000 549700 685000 364000 340000 ...
output	double [14 x 15557]	1.00 1.00 1.00 0.00 1.00 2.00 1.00 34.78 1.00 0.00 ...
dmunames	character [15557]	'H1' 'H2' 'H3' 'H4' 'H5' 'H6' ...
nc_inputs	NULL	Pairlist of length 0
nc_outputs	NULL	Pairlist of length 0
nd_inputs	NULL	Pairlist of length 0
nd_outputs	NULL	Pairlist of length 0
ud_inputs	NULL	Pairlist of length 0
ud_outputs	NULL	Pairlist of length 0

4.3.5 CLUSTER -5: (6,00,000 – 10,00,000)

Code -



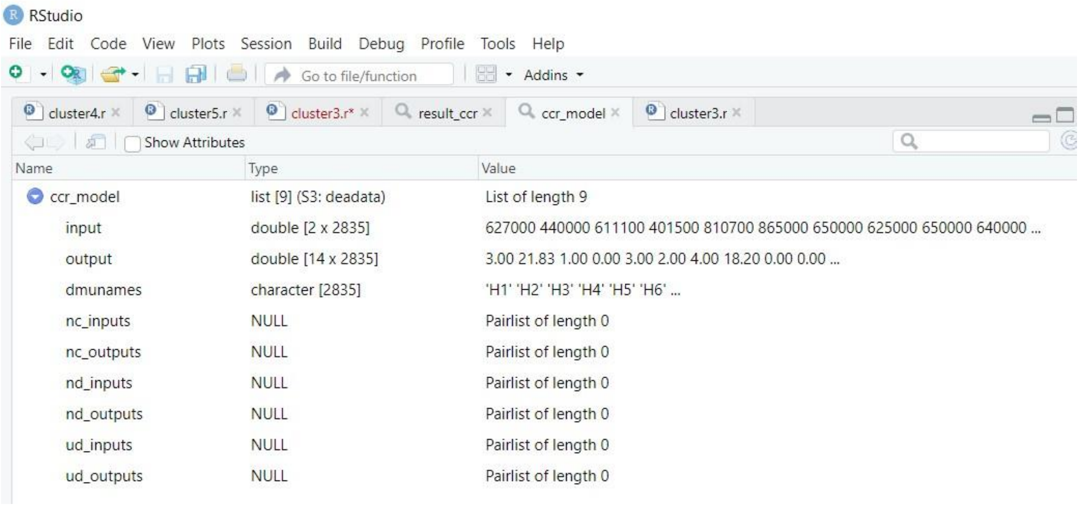

```
install.packages("deaR")
library(readxl)
data<-read_excel("C:\\Users\\HP\\Downloads\\(600000, g5).xlsx")
require(deaR)
ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
View(ccr_model)
result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
dmu_eval = 1:2835,dmu_ref = 1:2835)
efficiencies(result_ccr)
```

Output-

```
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/Ishika/
> data<-read_excel("C:\\Users\\HP\\Downloads\\(600000, g5).xlsx")
> require(deaR)
Loading required package: deaR
> ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
Warning messages:
1: 'read_data' is deprecated.
Use 'make_deadata' instead.
See help("Deprecated")
2: In make_deadata(datadea = datadea, ni = ni, no = no, dmus = dmus,  :
  There are negative or zero data. Try to translate the base point of the inputs/outputs with negati
ve data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
> result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs',
+ dmu_eval = 1:600,dmu_ref = 1:600)
> efficiencies(result_ccr)
       H1        H2        H3        H4        H5        H6        H7        H8        H9
0.9918379 1.0000000 0.8071518 0.9708514 0.9708514 0.9708514 0.9708514 1.0000000 0.9912489
      H10       H11       H12       H13       H14       H15       H16       H17       H18
0.9303082 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
      H19       H20       H21       H22       H23       H24       H25       H26       H27
1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9991897 0.9875260 1.0000000
      H28       H29       H30       H31       H32       H33       H34       H35       H36
1.0000000 0.9875260 1.0000000 0.9346552 0.9285260 0.9614560 0.9341660 0.9527388 0.9631922
      H37       H38       H39       H40       H41       H42       H43       H44       H45
1.0000000 0.7170154 0.9340079 1.0000000 1.0000000 1.0000000 0.9448192 0.9378492 0.9257426
      H46       H47       H48       H49       H50       H51       H52       H53       H54
0.9356223 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9848323 1.0000000
```

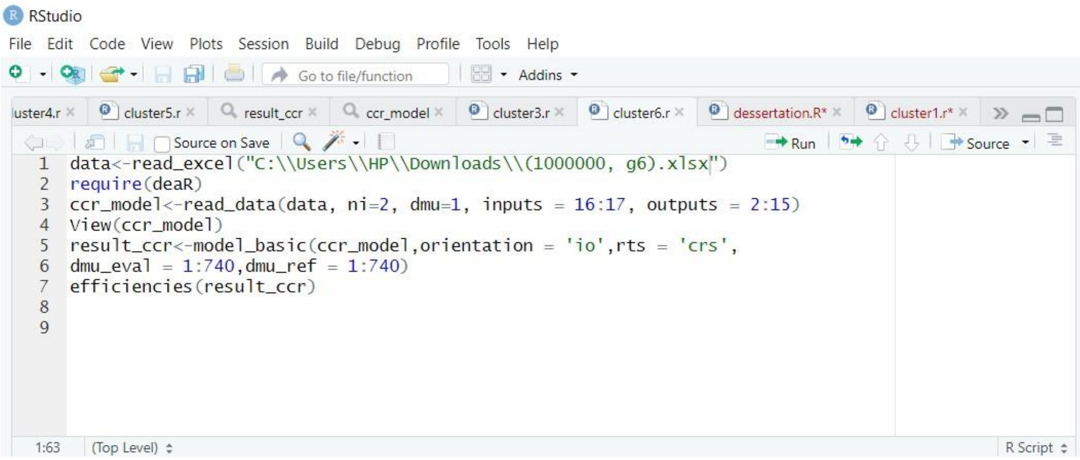
```
    H2753     H2754     H2755     H2756     H2757     H2758     H2759     H2760     H2761
1.0000000 1.0000000 0.9578022 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
    H2762     H2763     H2764     H2765     H2766     H2767     H2768     H2769     H2770
1.0000000 1.0000000 0.9891393 0.9455513 1.0000000 1.0000000 0.9736964 1.0000000 1.0000000
    H2771     H2772     H2773     H2774     H2775     H2776     H2777     H2778     H2779
0.9423742 1.0000000 1.0000000 1.0000000 1.0000000 0.9609042 1.0000000 1.0000000 1.0000000
    H2780     H2781     H2782     H2783     H2784     H2785     H2786     H2787     H2788
0.9493756 0.9925884 1.0000000 1.0000000 0.9501606 1.0000000 1.0000000 0.9517219 1.0000000
    H2789     H2790     H2791     H2792     H2793     H2794     H2795     H2796     H2797
1.0000000 0.9716636 0.9514668 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 0.9514668
    H2798     H2799     H2800     H2801     H2802     H2803     H2804     H2805     H2806
0.9725253 0.9791440 1.0000000 0.9663618 0.9548283 1.0000000 1.0000000 1.0000000 1.0000000
    H2807     H2808     H2809     H2810     H2811     H2812     H2813     H2814     H2815
1.0000000 1.0000000 1.0000000 0.9496982 0.9770269 1.0000000 1.0000000 0.9464030 1.0000000
    H2816     H2817     H2818     H2819     H2820     H2821     H2822     H2823     H2824
1.0000000 1.0000000 0.9751363 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
    H2825     H2826     H2827     H2828     H2829     H2830     H2831     H2832     H2833
1.0000000 1.0000000 1.0000000 0.8209134 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
    H2834     H2835
1.0000000 1.0000000
> 
```

View Model-



4.3.6 CLUSTER -6: (10,00,000 and above)

Code -

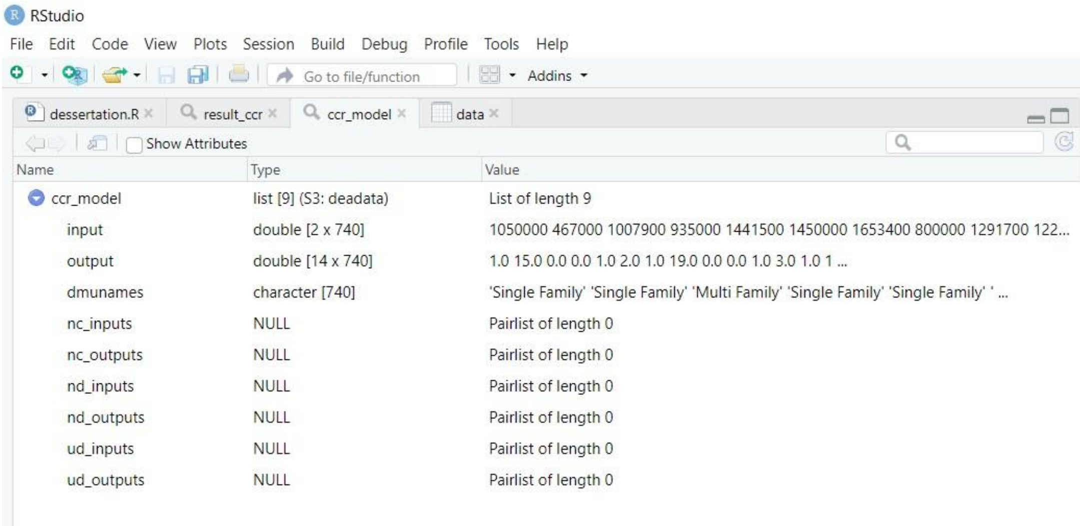


<u>OUTPUT-</u>

```
9:1    (Top Level) ÷                                                    R Script ÷
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/
> library(readxl)
> data<-read_excel("C:\Users\HP\Desktop\RStudio.xlsx")
> require(deaR)
> ccr_model<-read_data(data, ni=2, dmu=1, inputs = 16:17, outputs = 2:15)
Warning messages:
1: 'read_data' is deprecated.
Use 'make_deadata' instead.
See help("Deprecated")
2: In make_deadata(datadea = datadea, ni = ni, no = no, dmus = dmus,  :
  There are negative or zero data. Try to translate the base point of the inputs/outputs with negat
ive data in order to get only positive values.
  (This is a warning, not an error.)
> View(ccr_model)
> result_ccr<-model_basic(ccr_model,orientation = 'io',rts = 'crs', dmu_eval = 1:740, dmu_ref = 1:7
40)
> efficiencies(result_ccr)
Single Family Single Family  Multi Family Single Family Single Family Single Family
    1.0000000     0.9993162     0.9291426     0.7674195     0.9687021     0.9993694
Single Family Single Family Single Family Single Family Single Family Single Family
    0.9517688     0.9362806     0.9362806     0.9636158     1.0000000     0.9863832
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     0.9613269     1.0000000     0.9792676     1.0000000     0.9880477
Single Family Single Family Single Family Single Family Single Family Single Family
    0.9792676     0.9994568     0.8652095     0.9779213     0.9814670     0.8959690
```

```
9:1    (Top Level) ÷                                                    R Script ÷
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/
Single Family Single Family Single Family Single Family Single Family Single Family
    0.8138021     1.0000000     1.0000000     1.0000000     0.9766173     0.9695658
Single Family Single Family Single Family Single Family Single Family Single Family
    0.9749874     1.0000000     1.0000000     1.0000000     0.9076736     0.8670912
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     0.6865076     0.6354399     0.8506154     0.8155016     0.6856963
Single Family Single Family Single Family Single Family Single Family Single Family
    0.8632066     0.5630221     0.9798820     1.0000000     1.0000000     1.0000000
Single Family Single Family Single Family Single Family Single Family Single Family
    0.9785110     0.9751902     1.0000000     1.0000000     0.9818321     0.9799797
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     1.0000000     0.9802309     1.0000000     0.9957575     1.0000000
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     1.0000000     0.9796601     0.9700538     1.0000000     0.9431100
Single Family Single Family Single Family Single Family Single Family Single Family
    0.9722102     1.0000000     1.0000000     0.9497262     1.0000000     1.0000000
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     1.0000000     0.9830763     0.8650733     1.0000000     0.9776364
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     0.9860836     0.9820000     1.0000000     0.9747701     1.0000000
Single Family Single Family Single Family Single Family Single Family Single Family
    1.0000000     0.9787545     1.0000000     1.0000000     1.0000000     1.0000000
Single Family Single Family
    1.0000000     1.0000000
> |
```

View Model-


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
dessertation.R
result_ccr
ccr_model
data
Show Attributes
Name
Type
Value
ccr_model
list [9] (S3: deadata)
List of length 9
input
double [2 x 740]
1050000 467000 1007900 935000 1441500 1450000 1653400 800000 1291700 122...
output
double [14 x 740]
1.0 15.0 0.0 0.0 1.0 2.0 1.0 19.0 0.0 0.0 1.0 3.0 1.0 1 ...
dmunames
character [740]
'Single Family' 'Single Family' 'Multi Family' 'Single Family' 'Single Family' ' ...
nc_inputs
NULL
Pairlist of length 0
nc_outputs
NULL
Pairlist of length 0
nd_inputs
NULL
Pairlist of length 0
nd_outputs
NULL
Pairlist of length 0
ud_inputs
NULL
Pairlist of length 0
ud_outputs
NULL
Pairlist of length 0

Chapter 5 : CONCLUSION

Utilizing the outcomes of the Data Envelopment Analysis (DEA) and the efficiency calculations mentioned earlier, we can streamline the process of selecting an ideal house for the buyer. Tailoring the recommendations to the desired location, we can present a curated list of houses, each ranked for maximum efficiency based on predefined criteria. This approach empowers buyers to choose a home that aligns with their preferences from a selection optimized for efficiency.

Efficiency metrics span from 0 to 1, where 1 signifies the highest level of efficiency. By offering customers a list of houses prioritized by efficiency metrics, we enable them to make well-informed decisions. A house with an efficiency of 1 is considered the most efficient compared to others. The list consists of houses with maximum efficiency, ensuring that buyers can select a residence that not only meets their specific requirements but also provides maximum features and value within their budget constraints. Buying a house with an efficiency of 1 signifies the most efficient deal compared to houses with efficiencies less than one, providing buyers with a clear benchmark for optimal decision-making in their home purchase.

By offering customers a list of houses prioritized by efficiency metrics, we enable them to make well-informed decisions. This personalized strategy simplifies the home selection process, allowing customers to secure a residence that not only meets their specific requirements but also provides maximum features and value within their budget constraints. Ultimately, this data-driven approach enhances customer satisfaction, facilitating the acquisition of a home that perfectly matches their preferences while ensuring an efficient and valuable investment.

REFERENCES

1. An analysis of production as an efficient combination of activities _ CiNii Research. (n.d.).

2. Debreu, G. (1951). The Coefficient of Resource Utilization THE COEFFICIENT OF RESOURCE UTILIZATION'. In Source: Econometrica (Vol. 19, Issue 3).

3. Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. In Management Science (Vol. 30, Issue 9).

4. Coelli, T. J. (1996). A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. In CEPA Working Paper No. 8/96.

5. Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). The Measurement of Efficiency of Production. In Springer Science & Business Media.

6. Hsu, S. H., & Hwang, H. (2013). A fuzzy DEA/AR approach to the non-efficiency evaluation problem. In Expert Systems with Applications (Vol. 40, Issue 3).

7. Kao, C., & Liu, S. T. (2000). Data envelopment analysis with missing data: An application to university libraries in Taiwan. In Journal of the Operational Research Society (Vol. 51, Issue 8).

8. Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. In European Journal of Operational Research (Vol. 130, Issue 3).

9. Zhu, J. (2003). Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver. In Springer Science & Business Media.

10. Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. In Springer Science & Business Media.

11. Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) – Thirty years on. In European Journal of Operational Research (Vol. 192, Issue 1).

12. Sherman, H. D., & Gold, F. (1985). Bank branch operating efficiency: Evaluation with Data Envelopment Analysis. In Journal of Banking & Finance (Vol. 9, Issue 2).

13. Hollingsworth, B. (1999). Using data envelopment analysis in the public sector. In International Journal of Public Sector Management (Vol. 12, Issue 4).

14. Harker, P. T., & Friesz, T. L. (1984). Bounding the resource utilization function of an activity network. In Operations Research (Vol. 32, Issue 6).

15. Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. In Journal of Econometrics (Vol. 46, Issues 1-2).

16. Taylor & Francis Online: Peer-reviewed Journals (tandfonline.com)

17. Youssef Er-Rays, Meriem M'dioud, Analyzing the Efficiency of Moroccan Hospital Network Regions via DEA and Tobit Regression: Assessing DEAP 2.1 Software versus Generative AI ChatGPT 3.5

18. dspace.dtu.ac.in:8080

Appendix

Data features:

• House Code | Objective Feature | house_code | string |

• Category Code Description | Objective Feature| category_code_description | string |

• Exterior Condition | Objective Feature | exterior_condition | int |

• Frontage | Objective Feature | frontage | float (m) |

• Garage Spaces | Objective Feature | garage_spaces | int |

• House Extension | Objective Feature | house_extension | float (m^2) |

• Interior Condition | Objective Feature | interior_condition | int |

• Market Value | Objective Feature | market_value | float ($) |

• Number of Bathrooms | Objective Feature | number_of_bathrooms | int |

• Number of Bedrooms | Objective Feature | number_of_bedrooms | int |

• Number of Rooms | Objective Feature | number_of_rooms | int |

• Number of Stories | Objective Feature | number_stories | int |

• Sale Price | Objective Feature | sale_price | float ($) |

• Taxable Building Value | Objective Feature | taxable_building | float ($) |

• Taxable Land Value | Objective Feature | taxable_land | float ($) |

• Total Area | Objective Feature | total_area | float (m^2) |

• Total Liveable Area | Objective Feature | total_liveable_area | float (m^2) |

• Year Built | Objective Feature | year_built | int (year) |

	House_id	exterior_condition	frontage	garage_spaces	house_extension	interior_condition	number_of_bathrooms	number_of_bedrooms
0	H1	4	16.29	0	0	4	1	2
1	H2	4	12.00	0	0	4	1	0
2	H3	4	12.33	0	0	4	1	0
3	H4	4	12.33	0	0	4	1	0
4	H5	4	12.50	0	0	4	1	0
5	H6	4	12.50	0	0	4	1	0
6	H7	4	12.58	0	0	4	1	0
7	H8	4	14.08	0	0	4	1	0
8	H9	4	14.77	0	0	4	1	0
9	H10	7	14.00	0	0	7	1	2

number_of_rooms	number_stories	taxable_building	taxable_land	total_area	total_livable_area	year_built	market_value	sale_price	category_code_description
5	2	108300	36100	86.49	810	1920	144400	230000	Single Family
5	3	114835	20265	192.00	576	1920	135100	3500000	Single Family
5	3	116450	20550	197.28	591	1920	137000	3500000	Single Family
5	3	116450	20550	197.28	591	1920	137000	3500000	Single Family
5	3	117470	20730	200.00	600	1920	138200	3500000	Single Family
5	3	117470	20730	200.00	600	1920	138200	3500000	Single Family
5	3	117810	20790	201.28	603	1920	138600	3500000	Single Family
5	3	125545	22155	225.28	675	1920	147700	3500000	Single Family
5	3	127245	22455	236.32	690	1920	149700	3500000	Single Family
4	3	43480	65220	322.00	672	1920	108700	125000	Single Family

Table I.1 : Overview of the Dataset

LIST OF PUBLICATIONS

ICRTEBM, Amity University Conference Brochure

CERTIFICATE

AMITY UNIVERSITY   Springer

3rd INTERNATIONAL CONFERENCE ON
RECENT TRENDS IN ENGINEERING, TECHNOLOGY AND BUSINESS MANAGEMENT (ICRTETBM 2024)
(Digitization Transformation and Business Operations)

Organized By **Amity International Business School**
Amity Centre for Interdisciplinary Research

24th INTERNATIONAL BUSINESS HORIZON
INBUSH ERA WORLD SUMMIT 2024

21st to 23rd February 2024

CERTIFICATE OF PARTICIPATION

This is to certify that Ishika Gupta, Himanshi

from Department of Applied Mathematics, Delhi Technological University, Delhi, India has participated in **3rd International Conference on Recent Trends in Engineering, Technology and Business Management (ICRTETBM 2024)** held from February 21st to 23rd February, 2024 at Amity University, Noida, India.

He/She also chaired a session / delivered a keynote / invited talk / presented ✓ a paper

Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate

Prof. (Dr.) Gurinder Singh Group Vice Chancellor General Chair, ICRTETBM 2024	**Prof. (Dr.) P.K. Kapur** Director, ACIDR Conference Chair, ICRTETBM 2024	**Dr. Atul Chauhan** President, Ritnand Balved Education Foundation & Chancellor, Amity University

Publication Under Special Issue

ICRTETBM-2024 Conference Full Paper Submission for Special Issue "IJSAEM" Inbox ×

icrtetbm <icrtetbm@amity.edu> Wed, 27 Mar, 11:08
to icrtetbm

Dear Author(s),

I am pleased to notify you that your paper has been chosen for possible publication in the Special Issue of the International Journal of System Assurance Engineering and Management, Springer "IJSAEM" (https://www.springer.com/journal/13198). If you are interested, you may submit your paper to the Special Issue: 'S.I.: Leveraging Computational Paradigms for System Performance' (https://link.springer.com/journal/13198/updates/25916926) via the 'IJSAEM Editorial Manager- EM' (https://www.editorialmanager.com/ijsa/default.aspx).

Please ensure that you follow the Journal's guidelines while preparing your manuscript. You can refer to the author's guide in the International Journal of System Assurance Engineering and Management.

For all future correspondences, please mention your Paper ID received from springer after paper submission. **The Special Issue is not included in the article type, but you can select it in the additional information tab.**

****Please submit your manuscript through the Journal's homepage till April 20, 2024.**

##Please note that this email does not guarantee publication in the IJSA, as it is subject to the Journal's peer-review process.

With Best Wishes,
Technical Committee
ICRTETBM-2024
Amity University, Noida
...

12 of 140

Your submission to International Journal of System Assurance Engineering and Management: Preprint Confirmation Inbox ×

In Review at International Journal of System Assurance Engineering and Management <noreply@springernature.com> Sat, 20 Apr, 14:25
to me

Dear Ms. ISHIKA GUPTA,

Thank you for opting in to *In Review* from Springer Nature when you submitted your manuscript "Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate" to International Journal of System Assurance Engineering and Management.

In Review is a free service that posts your manuscript publicly and permanently to the preprint server Research Square and lets you track the status of your manuscript during the peer review process with a peer review timeline.

All preprints posted as part of the *In Review* service are screened for preprint suitability, formatted, assigned a DOI, and publicly posted under a CC-BY 4.0 license as a preprint on Research Square.

Your preprint may be posted within the next few days, prior to the beginning of the peer review process, or have a later posting date due to editorial processes at the journal. However, if you would like to preprint right away, please log in to Research Square and select "Post my preprint".

If you did not intend to opt in to *In Review* and would like to cancel your preprint, please log in to Research Square to opt out as soon as possible. Once your preprint is posted, you will no longer be able to have it removed.

PLAGIARISM REPORT

11% Overall Similarity

Top sources found in the following databases:

- 9% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 5% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	**dspace.dtu.ac.in:8080** Internet	6%
2	**deazone.com** Internet	1%
3	**Harvard Medical School Masters Program on 2024-05-01** Submitted works	<1%
4	**Universiti Teknologi Malaysia on 2016-07-02** Submitted works	<1%
5	**O.B Olesen. "Some unsolved problems in data envelopment analysis: A...** Crossref	<1%
6	**Ataturk Universitesi on 2024-02-27** Submitted works	<1%
7	**mme2020.mendelu.cz** Internet	<1%
8	**esd-conference.com** Internet	<1%

Similarity Report

9	University of Edinburgh on 2024-04-02 Submitted works	<1%
10	journalofbusiness.org Internet	<1%
11	Fatih University on 2014-12-30 Submitted works	<1%
12	Benchmarking: An International Journal, Volume 20, Issue 1 (2013-05... Publication	<1%
13	Cardiff University on 2012-10-12 Submitted works	<1%
14	H. David Sherman, Joe Zhu. "Benchmarking with quality-adjusted DEA ... Crossref	<1%
15	tidley.nfh.uit.no Internet	<1%

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Dissertation **Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate**

Total Pages __________

Name of the Scholar

(1) **Ishika Gupta**

(2) **Himanshi**

Department **Department of Applied Mathematics**

This is to report that the above dissertation was scanned for similarity detection. Process and outcome is given below:

Software used: ______________________ Similarity Index: _______________________

Total Word Count: _______________________

Date:

Candidate's Signature **Signature of Supervisor**

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042. India

CERTIFICATE OF DISSERTATION SUBMISSION FOR EVALUATION

1. Name: Ishika Gupta and Himanshi

2. Roll No. : 2K22/MSCMAT/58 and 2K22/MSCMAT/59

3. Dissertation title: Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate

4. Degree for which the dissertation is submitted: MSc Mathematics

5. Faculty of the University to which the dissertation is submitted: Anjana Gupta

6. Dissertation Preparation Guide was referred to for preparing the dissertation.

YES ☐ NO ☐

7. Specifications regarding dissertation format have been closely followed.

YES ☐ NO ☐

8. The contents of the dissertation have been organized based on the guidelines

YES ☐ NO ☐

9. The dissertation has been prepared without resorting to plagiarism.

YES ☐ NO ☐

10. All sources used have been cited appropriately. YES ☐ NO ☐

11. The dissertation has not been submitted elsewhere for a degree YES ☐ NO ☐

12. Submitted 2 spiral bound copies plus one CD. YES ☐ NO ☐

(Signature of Candidate)

Name(s): Ishika Gupta and Himanshi

Roll No: 2K22/MSCMAT/58 and 2K22/MSCMAT/59

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE OF FINAL DISSERTATION SUBMISSION

1. Names: Ishika Gupta and Himanshi
2. Roll No.: 2K22/MSCMAT/58 and 2K22/MSCMAT/59
3. Dissertation title: Efficiency-Based Housing Allocation: Leveraging DEA with CCR Model for Enhanced Decision-Making in Real Estate
4. Degree for which the dissertation is submitted: M.Sc. Mathematics
5. Faculty of the University to which the dissertation is submitted: Prof. Anjana Gupta
6. Dissertation Preparation Guide was referred to for preparing the dissertation.

   YES ☐ NO ☐
7. Specifications regarding dissertation format have been closely followed.

   YES ☐ NO ☐
8. The contents of the dissertation have been organized based on the guidelines

   YES ☐ NO ☐
9. The dissertation has been prepared without resorting to plagiarism.

   YES ☐ NO ☐
10. All sources used have been cited appropriately. YES ☐ NO ☐
11. The dissertation has not been submitted elsewhere for a degree. YES ☐ NO ☐
12. All the corrections has been incorporated. YES ☐ NO ☐
13. Submitted 2 spiral bound copies plus one CD. YES ☐ NO ☐

(Signature of the Supervisor)

Names : Dr. Anjana Gupta

(Signature of Candidates)

Names : Ishika Gupta

Himanshi

Roll No.: 2K22/MSCMAT/58

2K22/MSCMAT/59