HOUSE PRICE PREDICATION

A Project Report

Submitted By

HASANJI MARIYAM Z.-(210303128701)
SHAIKH SHIFA I.-(210303105807)
KUNAL KUMAR -(200303125079)
PATEL SAGAR R. -(210303128702)

in Partial Fulfilment For the Award of

the Degree of

BACHELOR OF TECHNOLOGY

COMPUTER SCIENCE AND ENGINEERING

Under the Guidance of

Prof. YASSIR FAROOQUI

Assistant Professor

PARUL UNIVERSITY

VADODARA

October - 2023

PARUL UNIVERSITY

CERTIFICATE

This is to Certify that Project - 2 -Subject code 203105400 of 7th Semester entitled “House Price Prediction” of Group No. PUCSE_167 has been successfully completed by

- HASANJI MARIYAM Z. - 210303128701
- SHAIKH SHIFA I. - 210303105807
- KUNAL KUMAR – 200303125079
- PATEL SAGAR R. - 210303128702

under my guidance in partial fulfillment of the Bachelor of Technology (B.TECH) in Computer Science and Engineering of Parul University in Academic Year 2023- 2024.

Date of Submission :-

Prof. Yassir Farooqui,	**Dr. Amit Barve**
Project Guide	Head of Department,
	CSE, PIET
Prof. Yatin Shukla	Parul University
Project Coordinator:-	External Examiner

Acknowledgements

Firstly from the bottom of my heart, I would like to express my sincere thanks to my guide, Supervisor **YASSIR FAROOQUI**, Assistant Professor, Computer Science and Engineering, I would like to express my sincere gratitude to all those who have helped me to complete my final year project. I am especially grateful to my supervisor, **YASSIR FAROOQUI** Sir , for their guidance, support, and encouragement throughout the project. I have learned a great deal from them, and I am truly grateful for their mentorship.

I would also like to thank the Computer Science department at Parul Institute of Engineering for providing me with the resources and facilities I needed to complete my project. I am particularly grateful to the technical staff for their help with troubleshooting and using the Facilities.

I would also like to thank my fellow students and researchers in the research group for their feedback and support. I have enjoyed working with them, and I have learned a lot from them.

This project was a significant learning experience, and I am deeply appreciative of everyone who played a part in its successful completion. Your contributions have been invaluable, and I am grateful for the knowledge and skills I have gained throughout this journey.

Finally, I would like to thank my family and friends for their love and support throughout the project. I could not have done it without them.

Thank you all for being a part of my final year project.

HASANJI MARIYAM (210303128701)
SHAIKH SHIFA (210303105807)
KUNAL KUMAR (200303125079)
PATEL SAGAR (210303128702)

Abstract

The Olympic games are international sports events with more than 200 nations participating in various competitions. The Sportspersons from various countries participate in competitions and make their countries proud of their excellence in sports. The primary objective of this paper is to analyze the Olympic dataset using python to compare overall performance of countries and to evaluate the contribution of each country in the Olympics. These analyses will give deeper insight into the performance of countries in Olympics over the years and helps sportspersons to quickly analyze their own and the competitor's performance. In this paper, the exploratory data analysis techniques are used to provide comparison between performance of various countries and the contribution of each country in the Olympics. Visualization of Olympics dataset in many aspects provides the status of countries in Olympics and helps countries with poor performance to produce quality players and improve nation's performance in Olympics. Despite a lot of hard work, many countries or players are unable to perform well during the events and grab medals whereas there are many countries that perform very well in the event and secure many medals. An analysis needs to be done by each country to evaluate the previous statistics which will detect the mistakes which they have done previously and will also help them in future development. Visualization of the data over various factors will provide us with the statistical view of the various factors which lead to the evolution of the Olympic Games and Improvement in the performance of various Countries/Players over time. The primary objective of this Research paper is to analyze the large Olympic dataset using Exploratory Data Analysis to evaluate the evolution of the Olympic Games over the years.
Keywords: Olympic, Sports, Nations, Python, Dataset.

Table of Contents

List of Figures

Chapter 1

OVERVIEW OF THE COMPANY

1.1 HISTORY

Code Unnati is a corporate-to-citizen digital literacy and IT skills development initiative, launched in June 2017. It is a multi-stakeholder, multi-year program that is a collaborative effort between SAP, CSR wings of different corporate companies, and Government as well as non-profit organizations. This program is aligned with the Digital India and Skill India initiatives of the Government of India under which a collaboration has been undertaken with 31 institutes in Gujarat to impart skills to their students in the domain of Industry 4.0.

The Centres of Excellence established by Edunet Foundation in collaboration with SAP across the selected engineering colleges in Gujarat, Maharashtra and Delhi will offer skilling opportunities to the students on the top emerging tech skills which include programing on Artificial Intelligence (AI) and Machine Learning (ML), Internet of Things (IoT), Data Science and Industry Specific SAP tools like SAP ABAP Programming and Data Analytics Tools SAP Analytics Cloud. The program helps the student acquire the skills relevant to the current industry needs, and thereby gain a competitive edge in the job market

Edunet Foundation is a social enterprise which was founded in 2015 and focuses on bridging the academia-industry divide, enhancing student employability, promoting innovation and creating an entrepreneurial ecosystem in India. Working primarily with emerging technologies, and striving to leverage them to augment, upgrade the knowledge ecosystem and equip the beneficiaries to become contributors themselves, we work extensively to build a workforce with an IR 4.0 enabled career..

The organization has enjoyed Special Consultative Status with the Economic and Social Council (ECOSOC) of the United Nations since 2020. With a national footprint, EF's programs, online and instructor-led, benefit tens of thousands of learners every year.

Code Unnati for the Value-Added course will be offered to those students who completed the Advance Course of Code Unnati. This course will cover the Concept

of IoT Cyber security and SAP analytics Cloud with Practical Hand on based Experimental Learnings.

1.2 SCOPE OF WORK

1.2.1 Educational Programs

Foundation Course

The Foundation Course under Code Unnati Program will be offered to the second and pre-final students pursuing engineering and other technical degree courses. This course will cover the prerequisites required for Artificial Intelligence, Machine Learning, Data Analytics, Deep Learnings, Computer Vision Technologies of the Code Unnati Advance and Value-Added Course for the pre-final/final year students.

Advance Course

Code Unnati for the Advance course will be offered to the pre-final/final year students pursuing engineering and other technical degree courses. This course will cover the Advance Concept of Artificial Intelligence, Machine Learning, Data Analytics, Deep Learnings, Computer Vision with OpenVINO Toolkit and Internet of Things with Practical Hands-on based Experimental Learnings.

Value Added Course

Code Unnati for the Value-Added course will be offered to those students who completed the Advance Course of Code Unnati. This course will cover the Concept of IoT Cyber security and SAP analytics Cloud with Practical Hand on based Experimental Learnings.

1.2.2 Student Support

Mentorship

Establish a mentorship program connecting students with experienced professionals in their chosen domains. Foster a supportive environment for students, addressing their concerns and providing career guidance.

Personalized Learning

Maintain small class sizes to facilitate personalized attention. Implement strategies to cater to individual learning styles and pace.

1.2.3 Infrastructure and Technology

Students Lms Portal

Ensure the smooth functioning of the online learning platform. Implement technological enhancements to improve the overall user experience.

IT Support

Provide technical support to students and faculty to address any platform-related issues. Implement strategies to cater to individual learning styles and pace. Stay updated on technological advancements to enhance the overall learning environment

1.2.4 Practical Learning Initiatives

Capstone Project

Ensure projects align with course objectives and enhance students' practical skills.

Internship Program

Provide guidance and support to students during their internships to maximize learning outcomes.

1.2.5 Continuous Improvement

Feedback Mechanism

Use feedback to continuously improve course content, delivery methods, and overall educational experience.

Professional Development

Provide ongoing training and professional development opportunities for faculty to stay abreast of industry developments.

1.3 ORGANIZATION CHART

Founder and CEO : Melinda Doe

Senior vice president : Jonathan Reckford

Chief people officer : Amy Dunham Direct Marketing : Jane Doe Finance Administration : Alex Krell

1.4 CAPACITY OF THE ORGANIZATION

Edunet Foundation is an Indian non profit with pan India presence. It enjoys Special Consultative Status with the Economic and Social Council (ECOSOC) of the United Nations from 2020. Having

Partner with 1,000s of institutions Across all levels of education and skilling , 300,000+ beneficiaries Covering a myriad of demographics , 8 programs that target all kinds of learners In collaboration with corporates and Govt. Bodies . The organization's strength lies not just in numbers but in the personalized attention and commitment to quality education it extends to each student. The intimate setting facilitates a deep engagement between the faculty and the diverse student body, fostering an atmosphere where learning is not just a process but a personalized journey.

Chapter 2

OVERVIEW OF DIFFERENT DEPARTMENTS

2.1 DEPARTMENT WORK

2.1.1 Product Development Unit

The product development unit carries out extensive market research including competitor analysis, skills gap analysis, surveys and focus group discussions. The curriculum design team defines program frameworks, learning objectives, and project timelines based on this research. Instructional designers conceptualize hands-on projects and assessments to test students' understanding of concepts. Content creators develop textual explainers, presentations, videos, and multimedia content supplemented by code samples to teach the required skills interactively.

2.1.2 Technology Unit

The technology unit architects and implements the core cloud infrastructure, software platforms, tools, and systems required to deliver a seamless learning experience. Key responsibilities include:

- Designing and implementing the Code Unnati application architecture using optimal technologies like Node.js, ReactJS, Docker, and Kubernetes.
- Building custom learning IDEs integrated with collaboration tools like Slack and GitHub to enable project-based learning and teamwork.
- Developing APIs and integrations to connect NullClass with external applications.

2.1.3 Content creation Unit

An in-house team of subject matter experts that have developed tailor-made pieces for over 200+ universities, courses, and programs to achieve desired learning outcomes

- Developing detailed lesson plans that include instructional strategies, activities, and assessments.
- Designing and creating curriculum materials for various subjects and grade levels.
- Aligning content with educational standards and learning objectives.
- Ensuring lessons are engaging, interactive, and cater to different learning preference
- Staying current with best practices in education through continuous learning.

2.1.4 Program Management Unit

Overseeing the full program deployment cycle right from program design through orientation, execution, and delivery, the experienced program management team is responsible for monitoring, evaluating, and scheduling all elements of a program.

- Overseeing the full program deployment cycle right from program design through orientation, execution, and delivery, the experienced program management team is responsible for monitoring, evaluating, and scheduling all elements of a program.
- Establish clear goals and performance indicators to measure the program's success.
- Allocate and manage resources, including budget, staff, and equipment, to ensure efficient program implementation.
- Identify and engage with key stakeholders, including government entities, community organizations, and other partners.
- Conduct periodic reviews to assess the effectiveness of program strategies.

2.1.5 Student Support Unit

The student support unit focuses on enabling an exceptional learning experience by providing prompt technical assistance and mentoring. Key tasks include.

- Guiding students through account creation, course enrolment, project onboarding, and platform usage

- Providing 24x7 help with code bugs, errors, IDE usage, tool configuration via chat, email and phone support.
- Conducting interactive live mentoring sessions to advise on project planning, architecture, and code reviews.
- Motivating students through community building initiatives on Slack and social media.
- Offering career guidance with resume building, mock interviews, and campus placement linkages.
- Gathering student feedback and sharing with product teams to enhance the learning experience.
- Monitoring student progress and proactively reaching out to offer help.

2.2 MAJOR SERVICES PROVIDED BY SAP

2.2.1 Enterprise Resource Planning

SAP (Systems, Applications, and Products) is a leading provider of Enterprise Resource Planning (ERP) software, offering a comprehensive suite of business applications to support various organizational functions. SAP ERP is a robust and integrated software solution designed to streamline and optimize business processes across different departments within an organization. Key features and aspects of SAP ERP include:

Modules

SAP ERP comprises various modules, each catering to specific business functions such as finance, human resources, supply chain management, production planning, sales and distribution, and more. These modules work seamlessly together, allowing for centralized data management and real-time information sharing.

Integration

One of the strengths of SAP ERP is its ability to integrate diverse business processes. This integration promotes efficiency by reducing data silos and ensuring consistency across the organization. For example, information entered into the system for sales can seamlessly update inventory levels and trigger production planning.

Centralized Database

SAP ERP operates on a centralized database, providing a single source of truth for data. This enables accurate and timely reporting, analytics, and decision-making.

Customisation

SAP ERP is highly customizable to meet the specific needs of different industries and organizations. This adaptability allows businesses to tailor the ERP system to their unique workflows and requirements.

Scalability

SAP ERP is scalable, making it suitable for small businesses as well as large enterprises. Organizations can start with basic modules and expand their ERP system as their business grows.

Real Time Analytics

The system provides real-time analytics and reporting capabilities, offering insights into key performance indicators (KPIs) to facilitate informed decision-making.

Cloud Option

In addition to on-premises solutions, SAP offers cloud-based ERP solutions. This allows businesses to benefit from the flexibility and accessibility of cloud computing.

2.2.2 CRM and Customer Experience (CX) Solutions from SAP

SAP provides a range of Customer Relationship Management (CRM) and Customer Experience (CX) solutions to help businesses manage and enhance their interactions with customers. These solutions are designed to streamline processes, improve customer satisfaction, and drive overall business success. Here are some key aspects of SAP's CRM and CX offerings:

SAP Customer Experience (SAP CX)

Commerce: SAP CX Commerce provides businesses with a platform to create personalized and seamless online shopping experiences for customers. It covers areas such as product content management, order management, and customer service. **Sales:** SAP CX Sales helps organizations manage their sales processes efficiently. It includes features for lead management, opportunity tracking, and sales forecasting, enabling sales teams to close deals more effectively. **Service:** SAP CX Service focuses on delivering excellent customer service experiences. It includes tools for managing customer inquiries, service requests, and support tickets, promoting timely and effective issue resolution.

Marketing: SAP CX Marketing supports organizations in creating targeted and personalized marketing campaigns. It includes features for campaign management, customer segmentation, and marketing analytics.

SAP Customer Data Cloud

SAP Customer Data Cloud provides a comprehensive solution for managing customer identities, consents, and preferences. It helps organizations build trust with customers by ensuring compliance with data protection regulations and delivering transparency in data usage.

Integration with SAP S/4HANA

SAP's CRM and CX solutions are often integrated with SAP S/4HANA, the company's Enterprise Resource Planning (ERP) suite. This integration ensures a seamless flow of information between customer-facing processes and backend operations, leading to a holistic view of customer interactions.

Cloud-Based Solutions

Many of SAP's CRM and CX solutions are available in the cloud, allowing businesses to leverage the benefits of scalability, flexibility, and accessibility associated with cloud computing.

Analytics and Insights

SAP CRM and CX solutions provide robust analytics and reporting capabilities, enabling businesses to gain valuable insights into customer behavior, preferences, and trends. This data-driven approach empowers organizations to make informed decisions and continuously improve their customer engagement strategies.

2.2.3 EPR and Financial Management

SAP ERP (Enterprise Resource Planning)

SAP S/4HANA: SAP S/4HANA is SAP's next-generation ERP suite designed to provide businesses with real-time insights and capabilities. It integrates various business functions, including finance, sales, supply chain, manufacturing, and more. S/4HANA leverages in-memory computing to enable faster data processing and analytics.

Core Modules: SAP ERP includes core modules such as Finance (SAP Finance), Supply Chain Management (SAP SCM), Human Capital Management (SAP HCM), and others. These modules are designed to streamline and optimize business processes, providing a unified platform for managing resources across the organization.

Cloud and On-Premises Deployment: SAP S/4HANA can be deployed both in the cloud and on-premises, offering flexibility to businesses based on their preferences and requirements.

SAP Financial Management Services

SAP Financials (SAP S/4HANA Finance): Formerly known as SAP Simple Finance, this module within S/4HANA focuses specifically on financial management. It includes features for financial accounting, management accounting, financial planning, and financial reporting. The aim is to provide a comprehensive and unified financial management solution.

SAP Concur: SAP Concur is a cloud-based solution that helps businesses manage travel, expense, and invoice processes. It streamlines expense reporting, automated invoice processing, and provides visibility into spending, contributing to better financial management.

SAP Analytics Cloud: While not exclusively a financial management tool, SAP Analytics Cloud integrates with SAP ERP systems to provide advanced analytics and business intelligence. This enables financial professionals to analyze financial data, create interactive dashboards, and make data-driven decisions.

SAP Treasury Management: SAP offers solutions for treasury management to help organizations manage their financial risks, cash, and liquidity effectively. This includes functionalities for cash management, risk management, and financial instruments.

2.2.4 Supply Chain Management (SCM) software Solution

SAP offers a comprehensive Supply Chain Management (SCM) software solution that is part of its broader SAP S/4HANA suite. This SCM solution is designed to help businesses optimize and streamline their end-to-end supply chain processes, from planning to execution. Here are key aspects of SAP's SCM software:

SAP Integrated Business Planning (IBP)

Demand Planning: SAP IBP includes advanced demand planning capabilities, allowing organizations to forecast and plan demand more accurately. It considers factors such as historical sales data, market trends, and external influences to provide insights into future demand.

Supply Planning: The solution enables effective supply chain planning by considering various factors, including inventory levels, production capacities, and supplier constraints. It helps organizations balance supply and demand to optimize their operations.

Sales and Operations Planning (S and OP): SAP IBP facilitates collaborative S and OP processes, allowing different departments within an organization to align on key planning decisions. This integration ensures that the entire organization is working towards common goals.

SAP Ariba Supply Chain Collaboration

SAP Ariba is a cloud-based procurement and supply chain collaboration platform. It connects buyers and suppliers, facilitating collaborative processes such as order collaboration, invoice collaboration, and supplier risk management.

SAP Extended Warehouse Management (EWM)

SAP EWM is a component of SAP S/4HANA that focuses on optimizing warehouse operations. It provides features for managing inventory, warehouse tasks, and distribution processes. SAP EWM supports advanced capabilities like slotting, cross-docking, and labor management.

SAP Transportation Management (TM)

SAP TM is designed to optimize transportation processes. It helps organizations plan and execute transportation activities, manage carrier relationships, and enhance visibility into shipments. SAP TM supports various transportation scenarios, including domestic and international logistics.

SAP Global Track and Trace

This solution provides end-to-end visibility into the supply chain, allowing organizations to track products as they move through the entire logistics network. It helps improve traceability, comply with regulations, and manage product recalls more effectively.

SAP Fiori User Experience (UX)

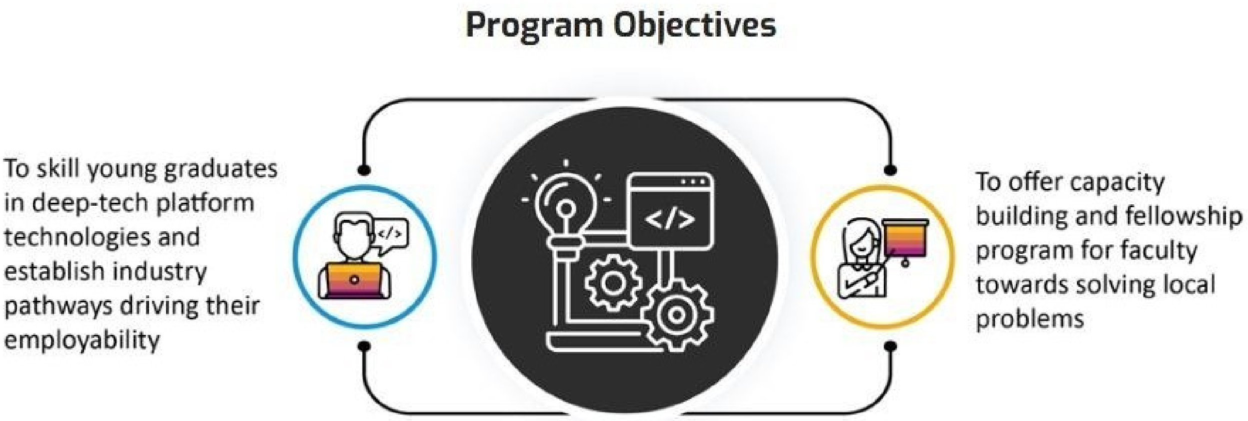
SAP Fiori provides a modern and intuitive user interface for SAP applications, including SCM. It enhances user experience, making it easier for users to interact with the system and access relevant information.

2.3 INFORMATION ABOUT THE PROGRAM

2.3.1 Program Objective

2.3.2 Key Features of the Program

2.3.3 Student-Centric Activities

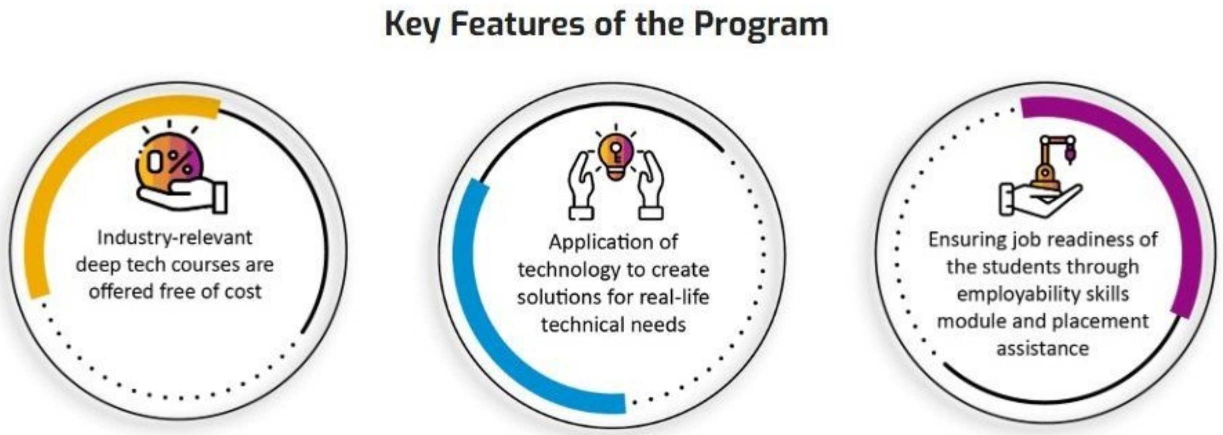


Program Objectives

To skill young graduates in deep-tech platform technologies and establish industry pathways driving their employability

To offer capacity building and fellowship program for faculty towards solving local problems

Figure 2.1: Enter Caption



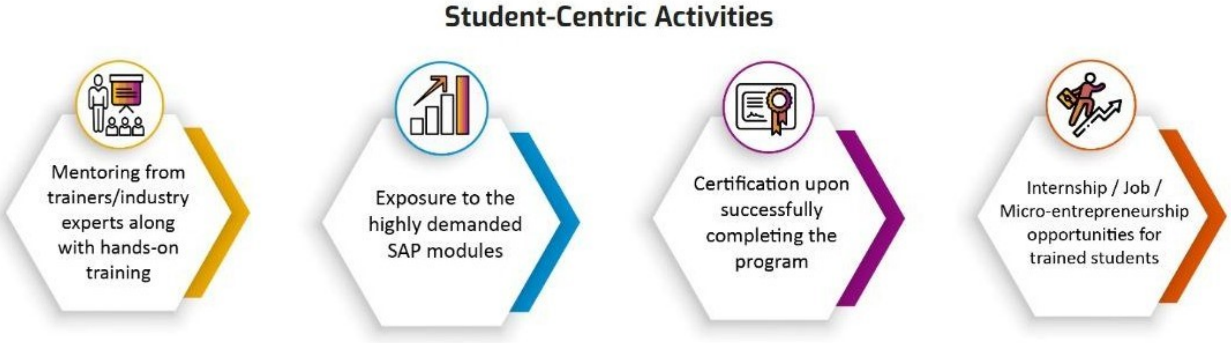
Key Features of the Program

Industry-relevant deep tech courses are offered free of cost

Application of technology to create solutions for real-life technical needs

Ensuring job readiness of the students through employability skills module and placement assistance

Figure 2.2: Enter Caption



Student-Centric Activities

Mentoring from trainers/industry experts along with hands-on training

Exposure to the highly demanded SAP modules

Certification upon successfully completing the program

Internship / Job / Micro-entrepreneurship opportunities for trained students

Figure 2.3: Enter Caption

Chapter 3

INTRODUCTION TO PROJECT

3.1 INTRODUCTION

Olympics is considered as the most important event worldwide, which provides a common platform to players from various nations to show their talents. The Olympics started in 1896, which is conducted once every four years. The goal of this paper is to analyze performance and participation of nations in Olympics from 1896 to 2016.In addition, the field of sports of particular country in particular year, in which they have contributed the maximum can be identified. The comparison of the performance of each sport with another can be done. The field of sports that has to have more participation can be identified and necessary action can be taken by players and nations to enhance themselves in future contributions towards the Olympics. The modern Olympic Games or Olympics are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are normally held every four years, alternating between the Summer and Winter Olympics every two years in the four years. Various scenarios come to our mind when we look into the Evolution of the Olympic Games over the years. These scenarios are: Increase in the number of participating nations, Increase in the number of participating Athletes, Increase/Decrease in the number of events, Increase in the expenditure cost of the event, improvement in the performance of the particular country, improvement in the performance of a particular player, Increase in women participation, Participation Ratio of Men to Women, improvement in medication facilities during competition, the effect of pandemic (if any) on the performance of the players. Analysis of these scenarios would depict the evolution of the Olympics over the years. This analysis would help in future prediction

3.2 METHODOLOGY

An Approach is referred to as a systematic path to reach a solution. Every problem, whether technical or non-technical, requires a proper approach so that we can get a proper path on which we have to proceed to get the required result. This Research Paper aims to analyze the vast history of Olympic Games and determine the evolution of Olympic Games over the Time. There are various factors which contribute to the evolution of the Olympics. To develop Olympics data analysis, we have followed methodology of:

3.2.1 Data Collection:

The very first step of any type of Analysis, whether it is technical or non-technical, is Data Collection. In order to perform analysis on a certain problem, we require a large amount of Data on which we apply various techniques and algorithms to reach a particular conclusion and get our desired result. It is advised to take the data in abundance because larger the volume of data for analysis, the greater would be the accuracy in the result and also the greater would be the confidence in decision making based on these results. We have used data from various data sources for analysis on Evolution of the Olympics over the time. We have taken three datasets which provide us with a large volume and a large variety of data for Analysis. The 1st dataset consists of information about the players and their entire details like their Gender, Height, Weight, Country for which they play, Medals won (Gold, Silver and Bronze) and many more. This data can be used to analyze the performance of the particular player and can also help in the comparative study between two or more players. the 2nd dataset consists of the information of the countries which have participated in the Olympics so far and the list of the total number of medals (Gold, Silver and Bronze) won by them. This data can be used to perform a comparative study on the performance of the countries. the 3rd dataset consists of the list of countries along with their country code which is the identification of these countries. This data can be used to find out the total number of countries which have participated in the Olympics so far.

3.2.2 Data Pre-Processing:

The next step after collecting Data is Data Processing. Data directly obtained from a data source such as dataset is known as Raw data. We can't apply various techniques or Machine Learning Algorithms like Linear Regression, Decision Tree, SVM etc directly to the Raw Data. This Data need to be processed and converted into useful data. Data Preprocessing is the process of translating the Raw data into Useful data by conscientiously checking for errors and eliminating redundant, incomplete, or incorrect data. The Dataset consists of various fields like Age, Gender, etc which consists of some null values which produces errors in the end result which is the Visualization of data in graphical format. These null values are needed to be omitted or replaced with some valid value which solves the error and generates accurate result. We have used a technique known as Deterministic Imputation to complete this task. Deterministic Imputation is a situation where the null values (NA or NaN) are determined with the help of the other values in the same column in the dataset. For this purpose, there are various models such as Basic Numeric Imputation Model in which the null value is replaced by Mean or Median of other values of the same column of the dataset. There is another model known as Hot Deck Imputation in which the null value is replaced by similar record in the dataset, i.e., some other value in the same column. Hot Deck Imputation can be applied to both Numerical as well as the Categorical value, but only if it contains enough values in the same column.

3.2.3 Exploratory Data Analysis:

The next step after data pre-processing is data analysis. In this step, analysis is done on data using various Techniques like Text Analysis, Diagnostic Analysis, Exploratory Data Analysis, etc and Machine learning Algorithms like Linear Regression, Logistic Regression, SVM, Decision Tree etc to reach to a particular conclusion. As our field of Research is visualization and comparative study of various factors which leads to the Evolution of Olympic games over the time, we are using the Exploratory Data Analysis technique to complete this task. Exploratory Data Analysis (EDA) is an approach to analyze data thoroughly and encapsulate its primary attributes basically in visual format. Exploratory Data Analysis is mainly used to see what the data represents apart from applying various algorithms. With the help of EDA, we can understand the structure and content of the dataset by various types of graphs and plots which can be drawn with the help of EDA. We can View the data in the visual format and can explain the analysis on that basis and perform a Comparative Study between different plots. There are various types of plots which are used in EDA. Some of them are mentioned below:

- Histogram
- Bar Graph
- Box Plot
- Scatter Plot and many more.

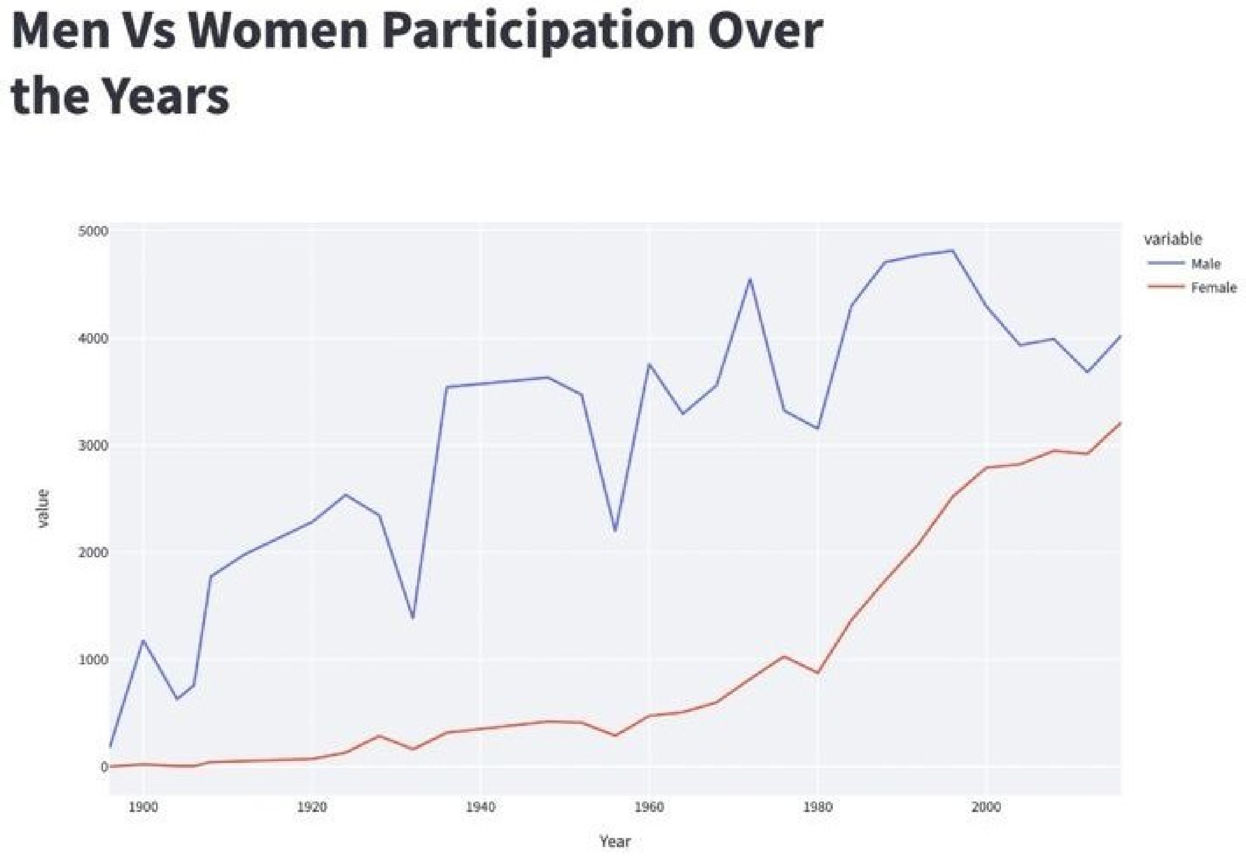


Figure 3.1: Enter Caption

3.3 ANALYSIS AND VISUALIZATION

1) Identifying Contribution of Men And Women Participants In Olympics (1896-2016)

The total number of men and women participants in the Olympics from 1896-2016 is analyzed and the ratio between men and women participants can be obtained. The analysis shows that the contribution of men is higher than women all over the world. The figure 1 shows gender wise contribution of players in the Olympics.

2) Identifying total number of medals achieved by USA Country in Olympics(1900-2000)

In this analysis, the total number of gold, silver and bronze medals won by the participants from all countries in the Olympics from 1896 to 2016 can be identified. The count includes the number of individuals who contributed separately or as a team to receive medals for their nations. The following results are obtained in the analysis. (i) USA has won the highest number of gold medals when compared to other medals and almost equal percent of silver and bronze medals. (ii)Australia has received the least number of gold medals when compared to other medals and won the highest number of bronze medals. Japan has fewer gold medals than other medals. France has less gold and a high number of silver and bronze medals.

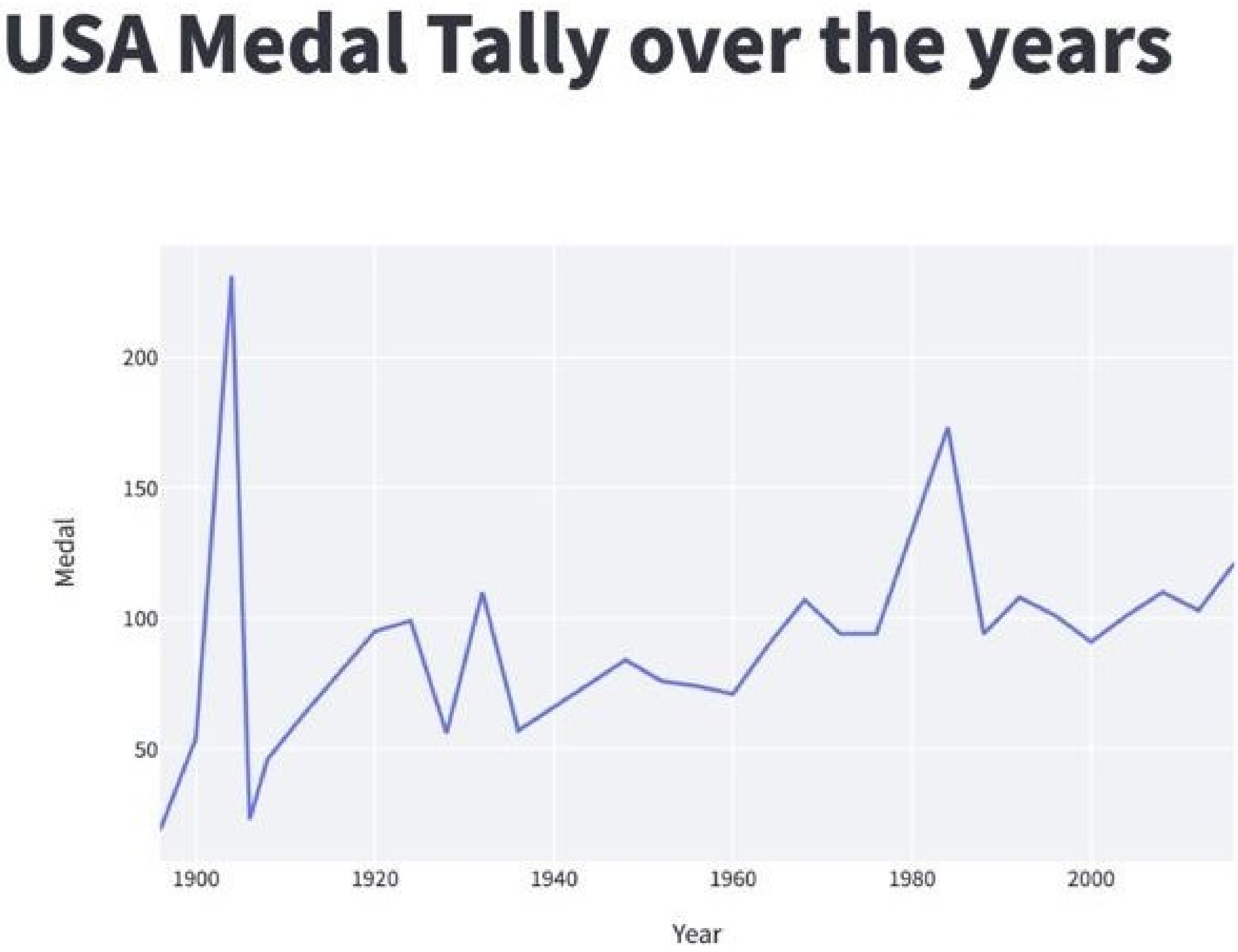


Figure 3.2: Enter Caption

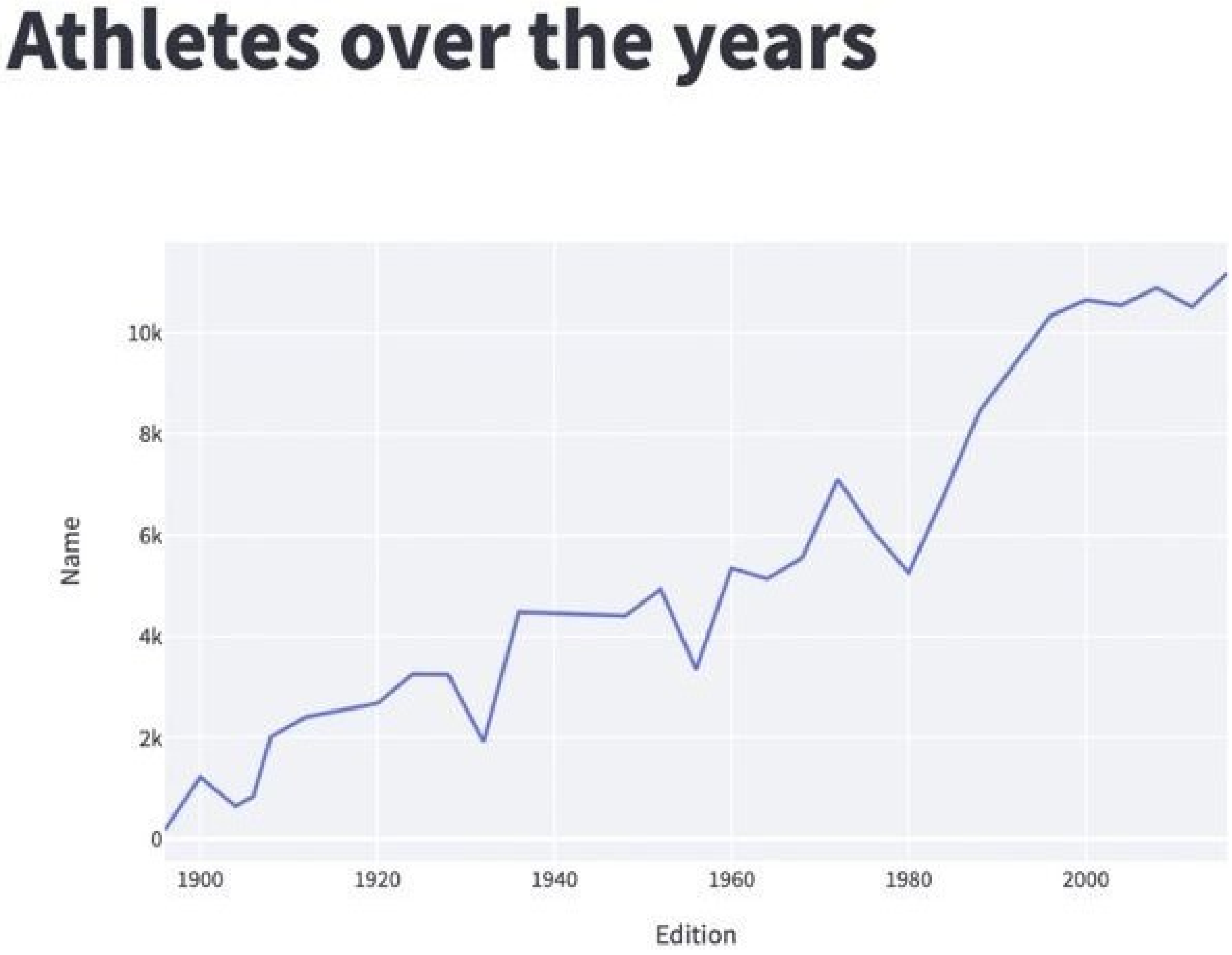
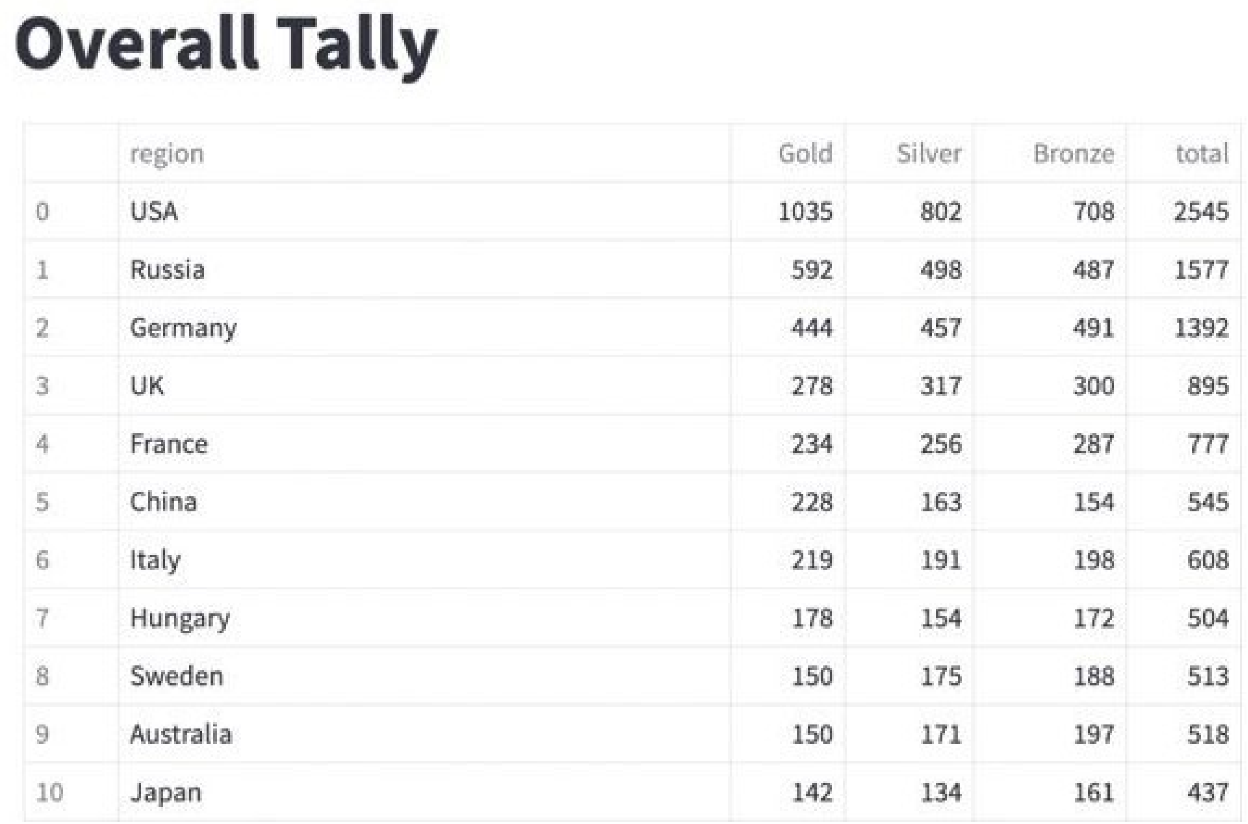


Figure 3.3: Enter Caption

3) Identifying the performance of Particular Country in Olympics (1992-2016)

Excellence of a country in the Olympics can be viewed by the number of medals won by a country. This analysis identifies the performance of a particular country in Olympics from 1992 to 2016.This can be processed by calculating the total medals won by a particular country in a particular year from 1992 to 2016. Data visualization can be carried out to represent the result of a particular country. The results are (i)Performance of India was gradually increasing from 1992 with no medals,1996 with 1 medal and finally in 2016 with 6 medals.(ii)Performance of USA was found like zig-zag graph from 1992 with 220 medals,1996 with 260 medals, suddenly performance has decreased in 2000 with 240 medals, increased gradually from 2004,contributed best in 2008 with 350 medals.(iii)Frances Performance was gradually increasing from 1996 to 2008 with medals within range of 40 and has performed well in 2016 with 80 medals.(iv)Performance of Australia was better during 1992 Olympics with 60 medals and there was a sudden increase in its performance with almost 200 medals over the period of 2000 and there has been gradual decrease in performance from 2004 to 2016.(v) Initially, performance of Japan was not so good ,but over the period of 2000



Overall Tally

	region	Gold	Silver	Bronze	total
0	USA	1035	802	708	2545
1	Russia	592	498	487	1577
2	Germany	444	457	491	1392
3	UK	278	317	300	895
4	France	234	256	287	777
5	China	228	163	154	545
6	Italy	219	191	198	608
7	Hungary	178	154	172	504
8	Sweden	150	175	188	513
9	Australia	150	171	197	518
10	Japan	142	134	161	437

Figure 3.4: Enter Caption

and 2004 there was a drastic increase in it and gained 100 medals which was higher than the rest.

4) Comparing the performance between the countries in Olympics(1896-2016)

The analysis compares the performance between the countries by medals won by the participants from selected countries in Olympics from 1896 to 2016.Countries such as USA, Hungary, France, Japan, Australia are selected for analysis .From this analysis, the following results has been inferred.(i)In the 1996 Olympics, among the five selected countries, USA is the leading country with a contribution of 7.53is the least country with 0.696.55among them.(iii)In 2004 Olympics, USA is the leading country with contribution of 8.053.3leading country with contribution of 8.52Hungary is the least country with 0.888.5among them.

5) Identifying the Best Performed Field of Sports for Particular Country in Olympics (2000-2016)

The analysis represents the performance from participants of a particular country and their best performing field of sport in Olympics from 2000 to 2016.To identify the field of sports of a particular country in a particular year and to analyze which field of sport has to have more participation. This provides information to enhance themselves in future contributions towards Olympics. a) In 2000, USA has performed best in the field of Aquatics and has performed least in the field of Weightlifting. b) In 2000, Australia has performed best in the field of Aquatics and has performed least in the field

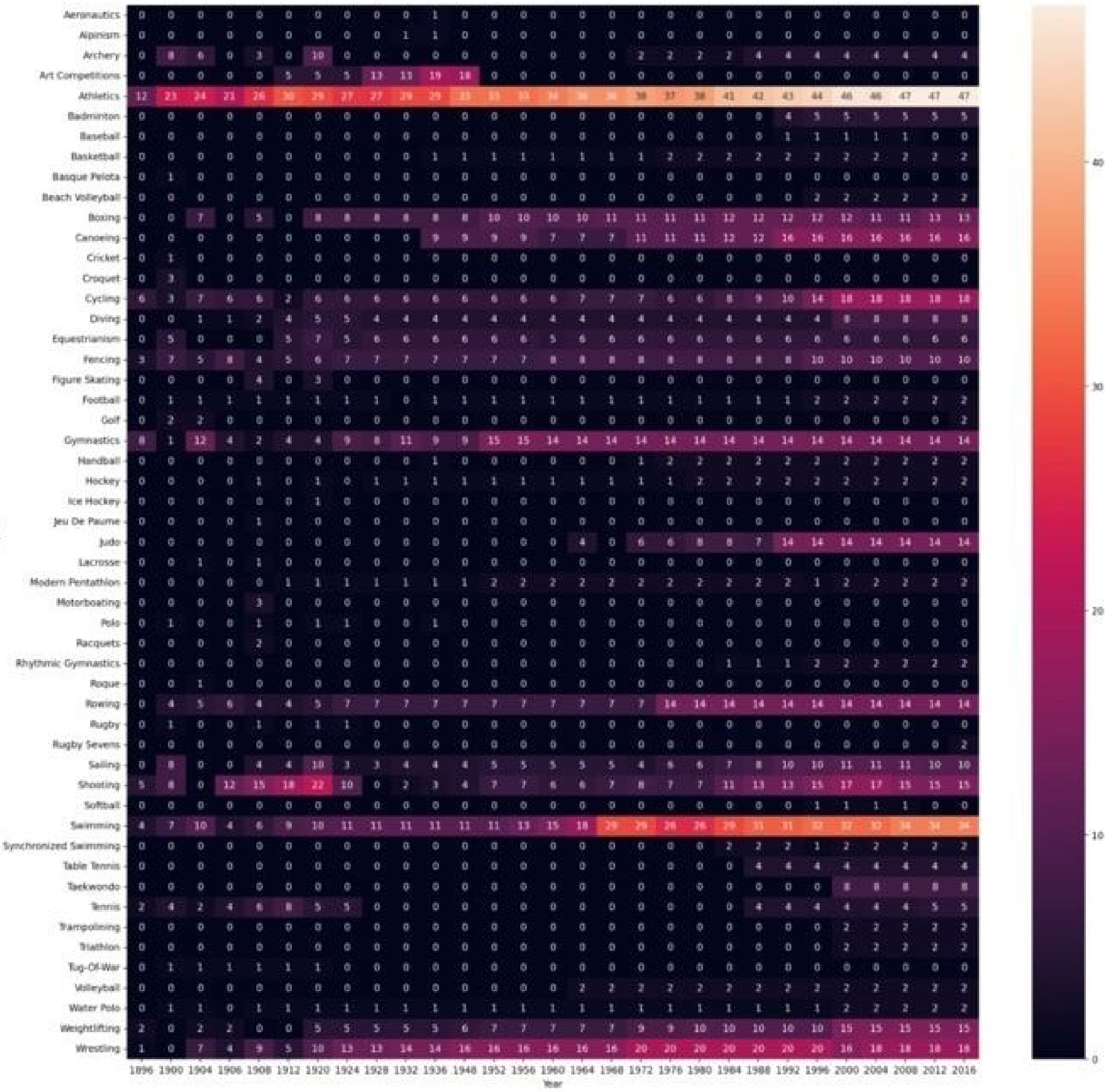
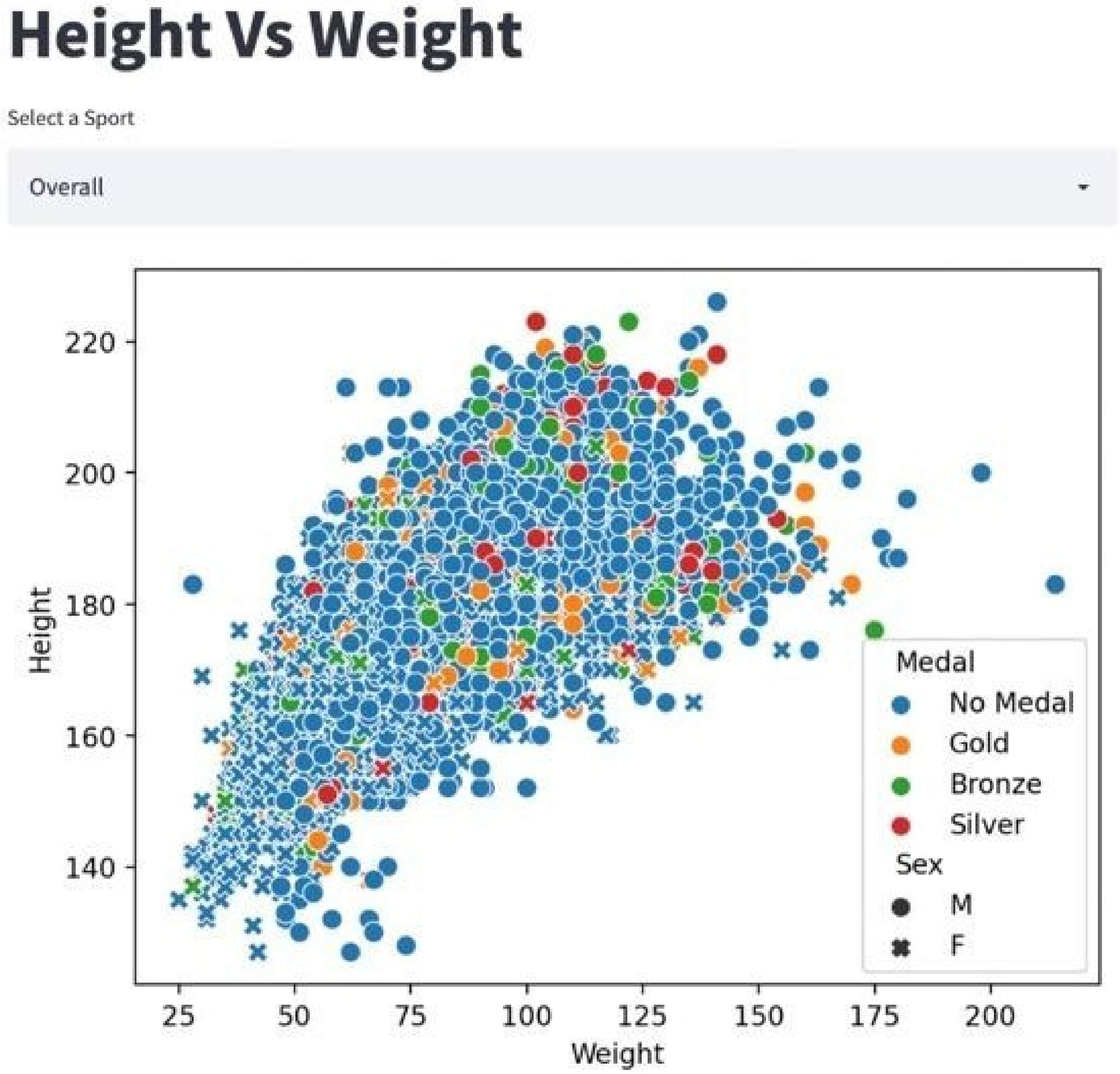


Figure 3.5: Enter Caption



Height Vs Weight

Select a Sport

Overall

Figure 3.6: Enter Caption

of Gymnastics. c) In 2000, France has performed best in the field of Fencing and has performed least in the field of Tennis. d) In 2000, Australia has performed best in the field of Aquatics and has performed least in the field of Athletics.

6) Analyzing the height vs weight

From this analysis, the Most Females who have won the medal are between 160-180cm tall and their weight class wide ranging from 50-150kg. The Number of Gold winners has performed well irrespective of their weight but seen density high at range of 175- 180cm height.

3.4 RESULT AND DISCUSSION

This work highlights the broad range of diagnostic and therapeutic services available to athletes during the London 2016 Olympic Games. Peak usage of many of the facilities was seen around days

9 and 10 of the competition (5 and 6 August 2016). This is when there was the greatest number of event finals occurring11 and the athletes' village was at its busiest. As expected, most consultations were musculoskeletal in origin but a sizable proportion also related to dental and ophthalmic complaints. The demand for MRI was significant, reflecting the fact that this resource is considered not as freely available as others. wise as it is during Games time. Pathology investigations were performed steadily throughout the period of competition, but the demand for pharmacy services did spike considerably. It is interesting to note from the continent subanalysis that the greatest proportion of attendances was from athletes from African nations. This was for the gross number of overall attendances and also when corrected for multiple attendances by individual athletes. It is also interesting that although Oceania provided the smallest proportion of overall attendances (7Oceania fielded the smallest number of athletes (670); therefore, individual attendances would constitute a greater proportion of the small Oceania cohort. Athletes were able to self-present to the Polyclinic and would often be accompanied by their NOC's medical or administrative staff. On arriving at the Polyclinic, they were quickly triaged to the appropriate department and rarely had a significant delay in being seen. Staffing levels appeared to meet the demands effectively; however, minimal waiting time was seen for some of the busier services such as physiotherapy, sports massage and radiology. Despite being serviced entirely by volunteers, staff had undergone a comprehensive recruitment and selection process involving an induction and orientation to the building and working environment prior to the start of the Games. This enabled an efficient working environment right from the start of the Games and limited any start-up issues. Daily work force meetings at the start and end of each shift further reinforced good communication and working relations among staff from different departments in the Polyclinic. Efficient assimilation and storage of medical encounter data were crucial throughout the Games. Workstations connected to the Games network were available in all medical venues including all fields of play to allow timely data input. This meant that records were kept contemporaneously and could be referred to during successive visits for the same individual. The Atos database provided an effective platform for these data to be securely stored and contained relevant data fields to be comprehensive and appropriate.

Chapter 4

Implementation

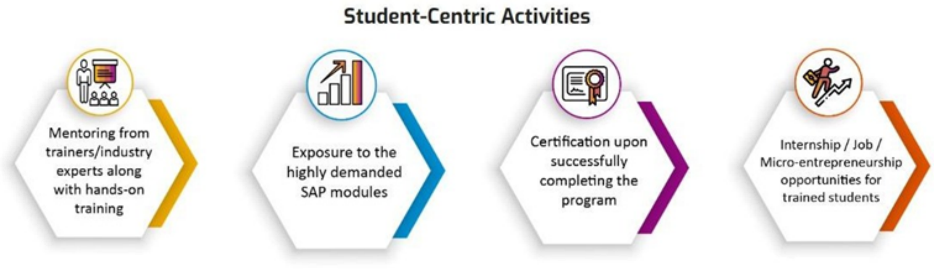
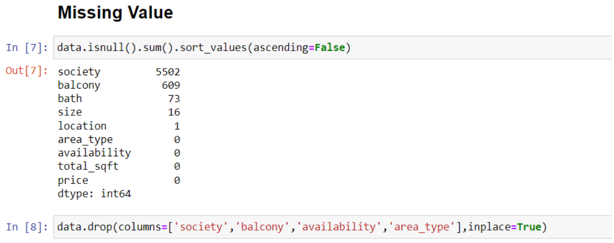

Student-Centric Activities
Mentoring from trainers/industry experts along with hands-on training
Exposure to the highly demanded SAP modules
Certification upon successfully completing the program
Internship / Job / Micro-entrepreneurship opportunities for trained students

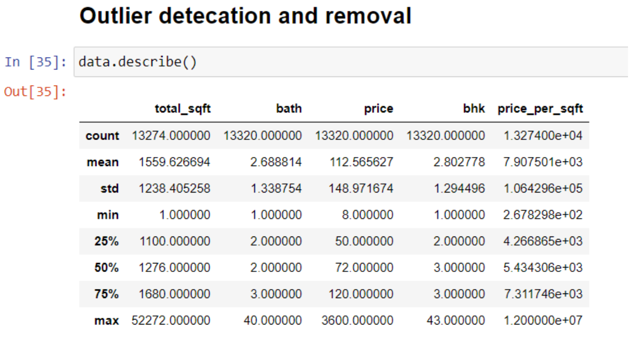
Figure 4.1: Import DataSet



Missing Value

```
In [7]: data.isnull().sum().sort_values(ascending=False)

Out[7]: society         5502
        balcony          609
        bath              73
        size              16
        location           1
        area_type          0
        availability       0
        total_sqft         0
        price              0
        dtype: int64

In [8]: data.drop(columns=['society','balcony','availability','area_type'],inplace=True)
```

Figure 4.2: handle Missing Value

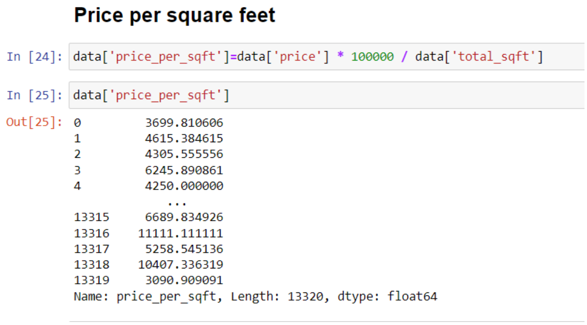


Outlier detecation and removal

```
In [35]: data.describe()

Out[35]:
```

	total_sqft	bath	price	bhk	price_per_sqft
count	13274.000000	13320.000000	13320.000000	13320.000000	1.327400e+04
mean	1559.626694	2.688814	112.565627	2.802778	7.907501e+03
std	1238.405258	1.338754	148.971674	1.294496	1.064296e+05
min	1.000000	1.000000	8.000000	1.000000	2.678298e+02
25%	1100.000000	2.000000	50.000000	2.000000	4.266865e+03
50%	1276.000000	2.000000	72.000000	3.000000	5.434306e+03
75%	1680.000000	3.000000	120.000000	3.000000	7.311746e+03
max	52272.000000	40.000000	3600.000000	43.000000	1.200000e+07

Figure 4.3: Outlier detection and removel



Price per square feet

```
In [24]: data['price_per_sqft']=data['price'] * 100000 / data['total_sqft']

In [25]: data['price_per_sqft']

Out[25]: 0         3699.810606
         1         4615.384615
         2         4305.555556
         3         6245.890861
         4         4250.000000
                      ...
         13315     6689.834926
         13316    11111.111111
         13317     5258.545136
         13318    10407.336319
         13319     3090.909091
         Name: price_per_sqft, Length: 13320, dtype: float64
```

Figure 4.4: Feature selection

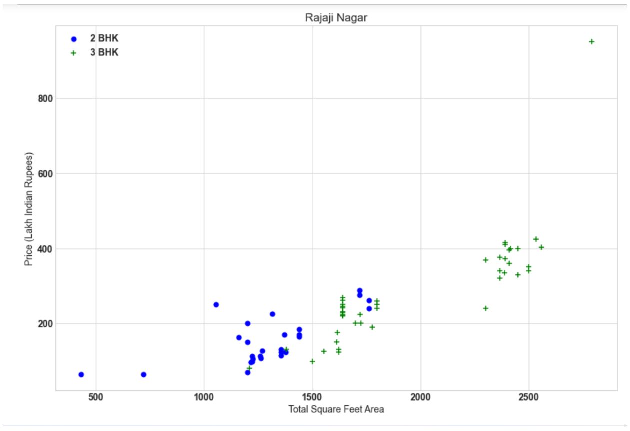


Figure 4.5: Total Square Feet area base on price

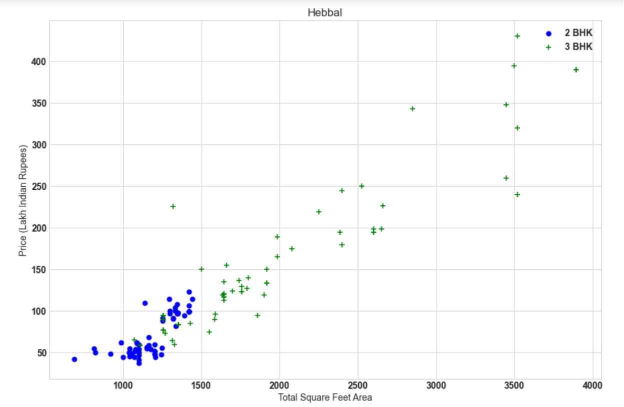


Figure 4.6: Total Square Feet area base on price

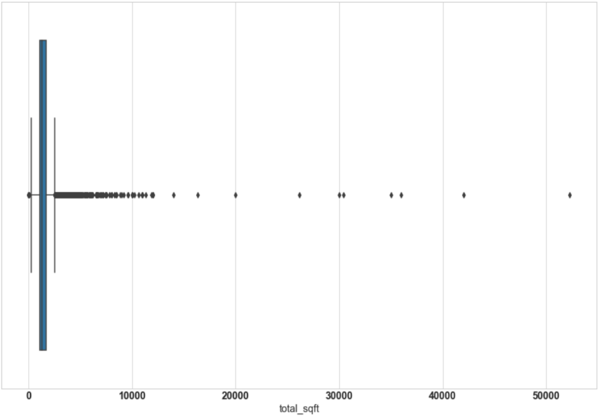


Figure 4.7: Total Square feet

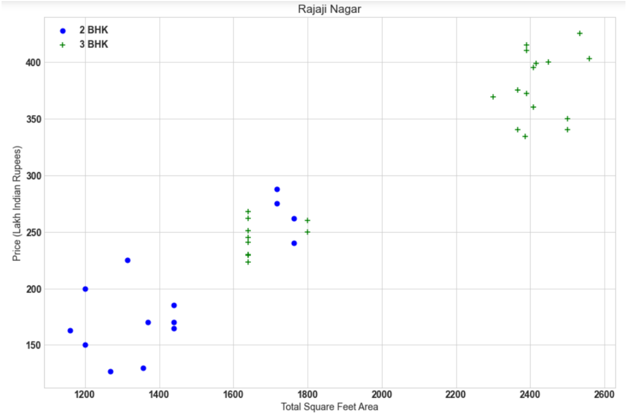


Figure 4.8: Total Square Feet area base on price

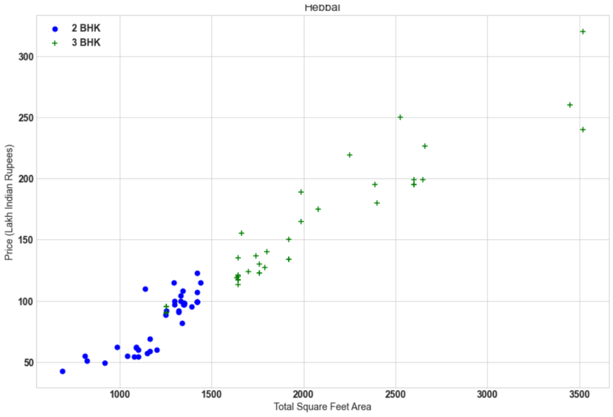


Figure 4.9: Total Square Feet area base on price

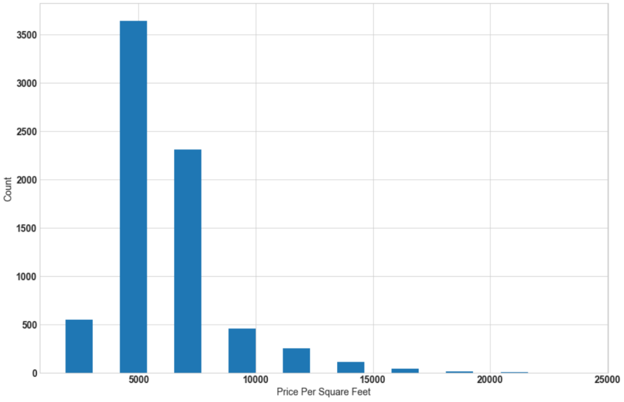


Figure 4.10: Enter Caption

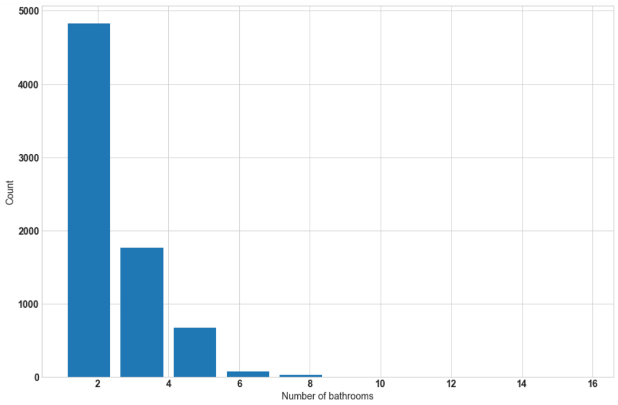


Figure 4.11: Price per squre feet based on count

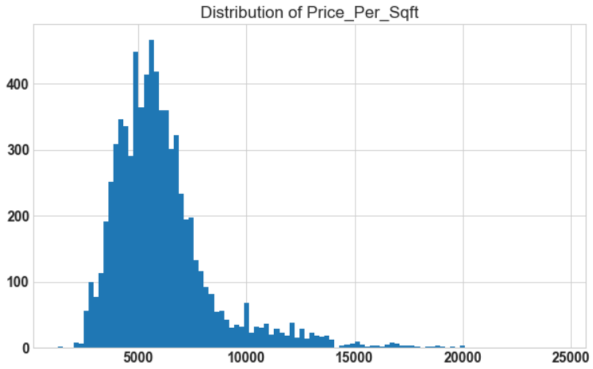


Figure 4.12: Distribution of price squre feet

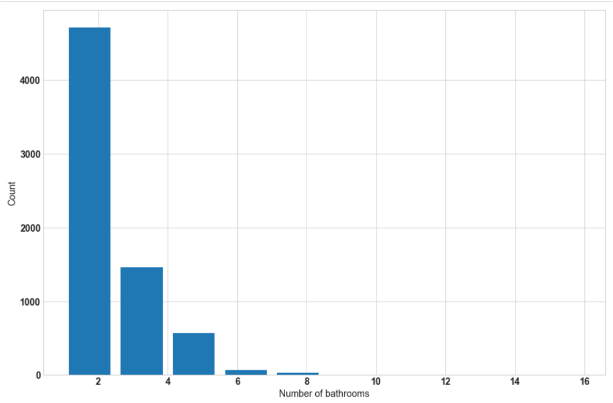
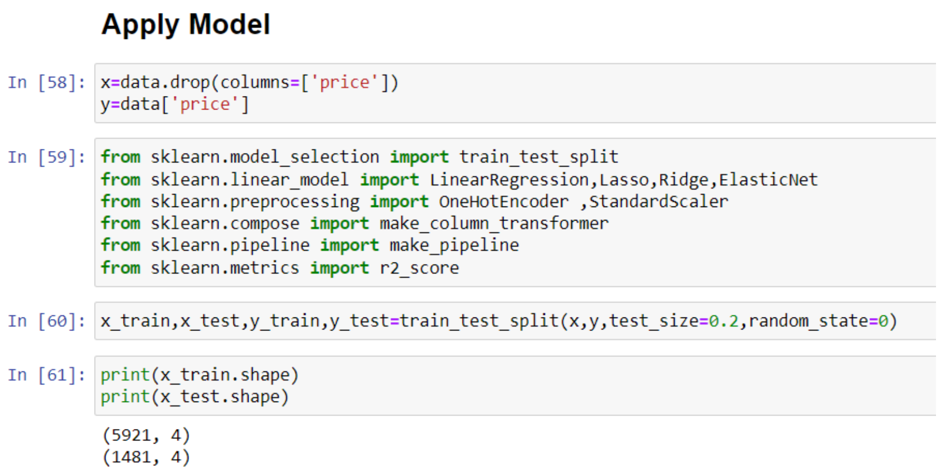


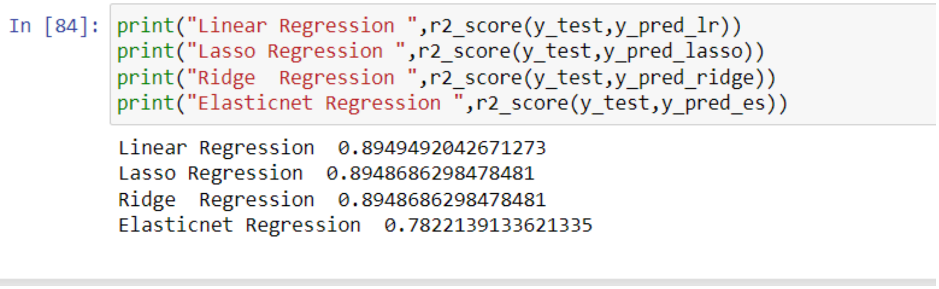
Figure 4.13: No. of bathroom count



Apply Model

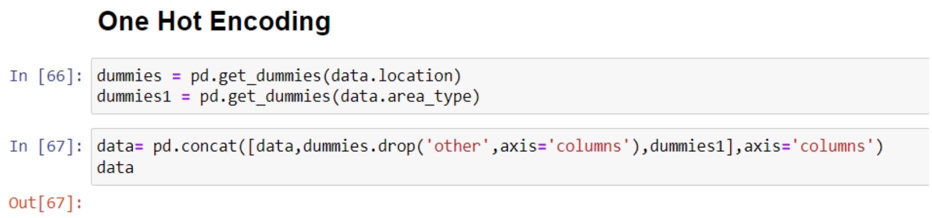
```
In [58]: x=data.drop(columns=['price'])
         y=data['price']

In [59]: from sklearn.model_selection import train_test_split
         from sklearn.linear_model import LinearRegression,Lasso,Ridge,ElasticNet
         from sklearn.preprocessing import OneHotEncoder ,StandardScaler
         from sklearn.compose import make_column_transformer
         from sklearn.pipeline import make_pipeline
         from sklearn.metrics import r2_score

In [60]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)

In [61]: print(x_train.shape)
         print(x_test.shape)

         (5921, 4)
         (1481, 4)
```

Figure 4.14: Apply Model



```
In [84]: print("Linear Regression ",r2_score(y_test,y_pred_lr))
         print("Lasso Regression ",r2_score(y_test,y_pred_lasso))
         print("Ridge  Regression ",r2_score(y_test,y_pred_ridge))
         print("Elasticnet Regression ",r2_score(y_test,y_pred_es))

         Linear Regression  0.8949492042671273
         Lasso Regression  0.8948686298478481
         Ridge  Regression  0.8948686298478481
         Elasticnet Regression  0.7822139133621335
```

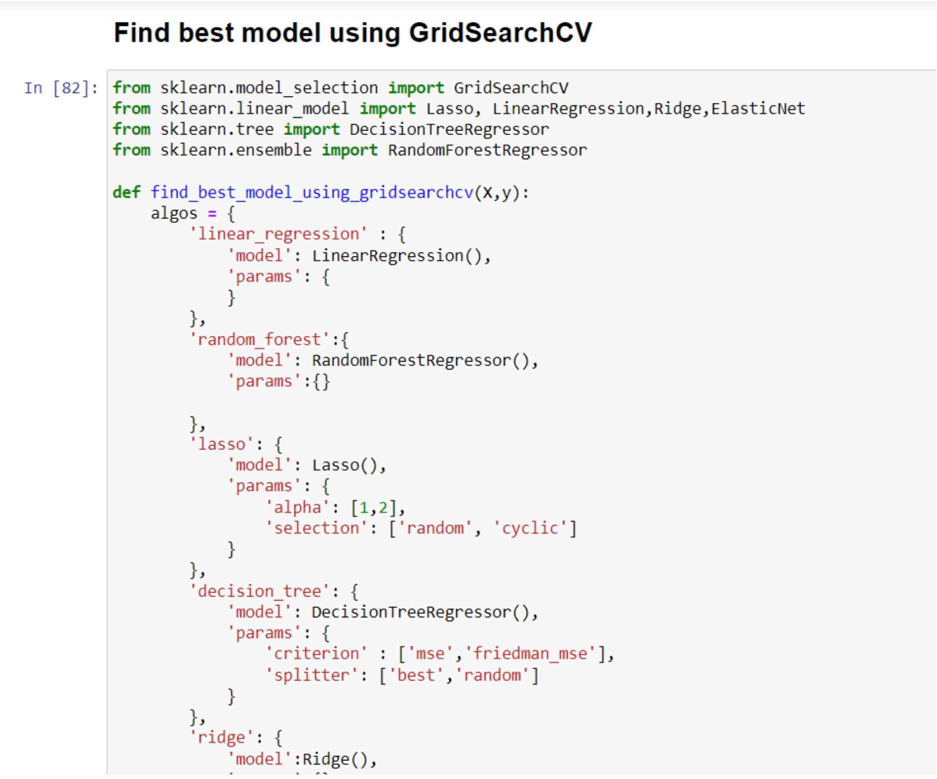
Figure 4.15: Model accuracy



One Hot Encoding

```
In [66]: dummies = pd.get_dummies(data.location)
         dummies1 = pd.get_dummies(data.area_type)

In [67]: data= pd.concat([data,dummies.drop('other',axis='columns'),dummies1],axis='columns')
         data

Out[67]:
```

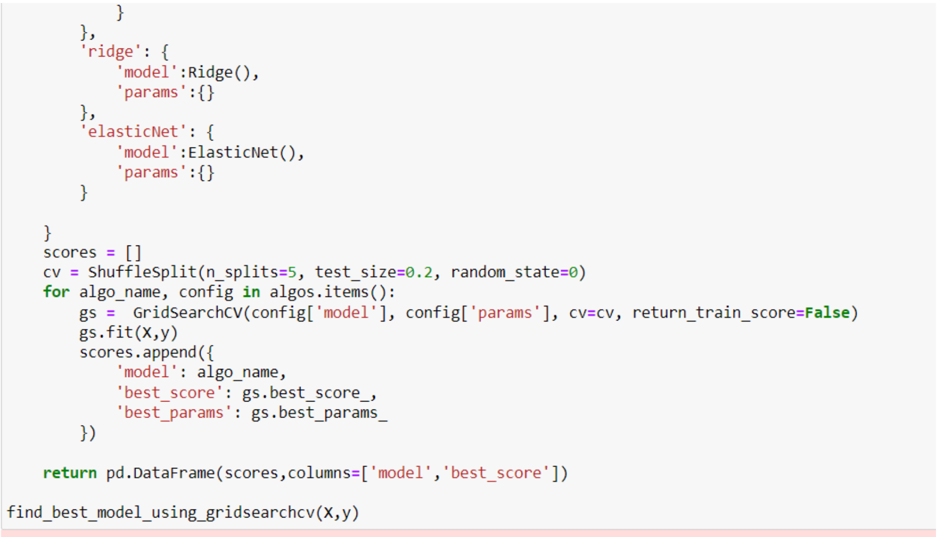
Figure 4.16: One Hot Encoding



Find best model using GridSearchCV

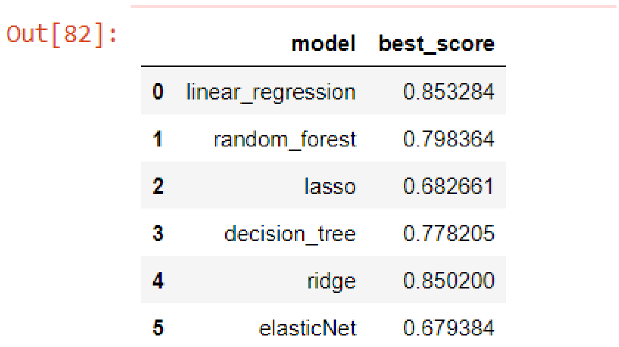
```
In [82]: from sklearn.model_selection import GridSearchCV
         from sklearn.linear_model import Lasso, LinearRegression,Ridge,ElasticNet
         from sklearn.tree import DecisionTreeRegressor
         from sklearn.ensemble import RandomForestRegressor

         def find_best_model_using_gridsearchcv(X,y):
             algos = {
                 'linear_regression' : {
                     'model': LinearRegression(),
                     'params': {
                     }
                 },
                 'random_forest':{
                     'model': RandomForestRegressor(),
                     'params':{}

                 },
                 'lasso': {
                     'model': Lasso(),
                     'params': {
                         'alpha': [1,2],
                         'selection': ['random', 'cyclic']
                     }
                 },
                 'decision_tree': {
                     'model': DecisionTreeRegressor(),
                     'params': {
                         'criterion' : ['mse','friedman_mse'],
                         'splitter': ['best','random']
                     }
                 },
                 'ridge': {
                     'model':Ridge(),
```

Figure 4.17: Apply all the model



```
            }
        },
        'ridge': {
            'model':Ridge(),
            'params':{}
        },
        'elasticNet': {
            'model':ElasticNet(),
            'params':{}
        }

    }
    scores = []
    cv = ShuffleSplit(n_splits=5, test_size=0.2, random_state=0)
    for algo_name, config in algos.items():
        gs =  GridSearchCV(config['model'], config['params'], cv=cv, return_train_score=False)
        gs.fit(X,y)
        scores.append({
            'model': algo_name,
            'best_score': gs.best_score_,
            'best_params': gs.best_params_
        })

    return pd.DataFrame(scores,columns=['model','best_score'])

find_best_model_using_gridsearchcv(X,y)
```

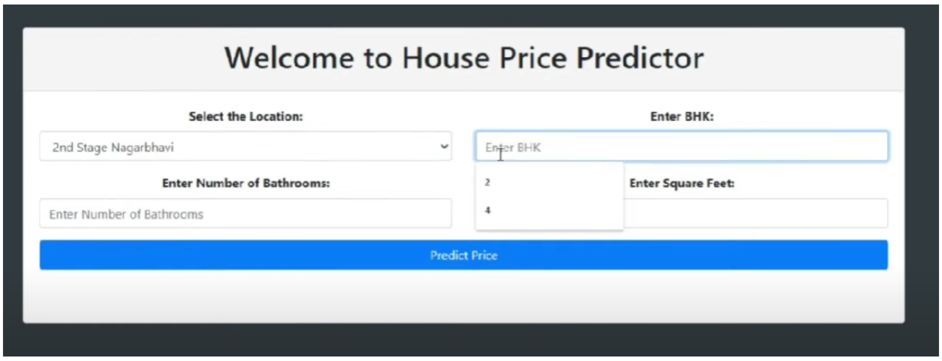
Figure 4.18: Apply all the model



Out[82]:

	model	best_score
0	linear_regression	0.853284
1	random_forest	0.798364
2	lasso	0.682661
3	decision_tree	0.778205
4	ridge	0.850200
5	elasticNet	0.679384

Figure 4.19: Accuracy of the model



Welcome to House Price Predictor

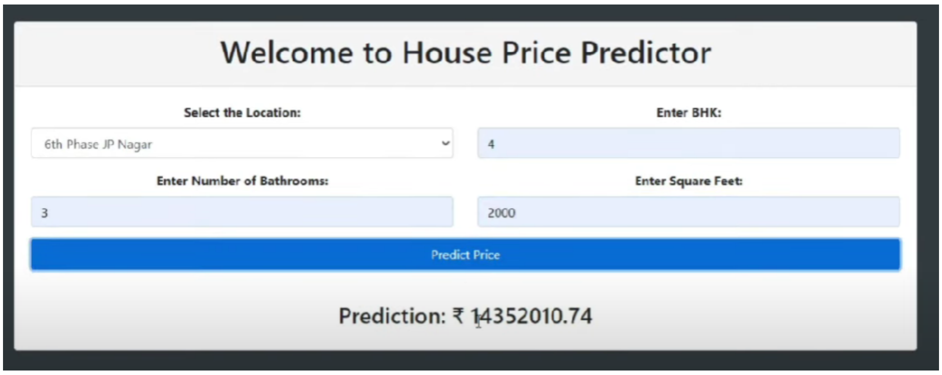
Select the Location: 2nd Stage Nagarbhavi

Enter BHK: Enter BHK

Enter Number of Bathrooms: Enter Number of Bathrooms

Enter Square Feet:

Figure 4.20: Home Page



Welcome to House Price Predictor

Select the Location: 6th Phase JP Nagar

Enter BHK: 4

Enter Number of Bathrooms: 3

Enter Square Feet: 2000

Prediction: ₹ 14352010.74

Figure 4.21: Home Page

Chapter 5

Conclusion

5.1 Conclusion:

The main objective of this study was to analyze and visualize the various factors which have contributed to the Evolution of the Olympic Games over the years. This type of analysis is very helpful as this type of analysis can be performed by any Country or Player which can help them in analyzing their performance so that they can improve their performance by changing their strategies. We have used a technique named Exploratory Data Analysis which enables you to encapsulate the primary factors of a dataset into a visual format. We selected Python language to implement our work because It is one of the best languages suitable for Data Analysis and is the platform where we have performed this Analysis. As a result of the Analysis, we can conclude that It is true that Olympic Games have evolved considerably over time from the 1896 Olympic Games till the 2016 Rio Olympics. Various factors provide valid evidence that the Olympics have changed a lot. some of these factors are the launch of the Winter Olympic Games apart from the Summer Olympic Games in 1924, an increase in the number of participating countries in both Summer and Winter Olympics, the Average age of players in the Olympic Games, the increase in the participation of the females in both Summer and Winter Olympics over the time, Total number of medals won by various participating countries over the years, Average height and the weight of Players who contributes to victory of Games in the event. Apart from these, there are many more factors that depict the Evolution of the Olympic Games over time. Visualization of these factors has been done to explain and validate the Analysis in various Graphical formats like a Line graph, Scatter Plots, Bar, Graphs, Dist Plots, etc.

Chapter 6

Future Work

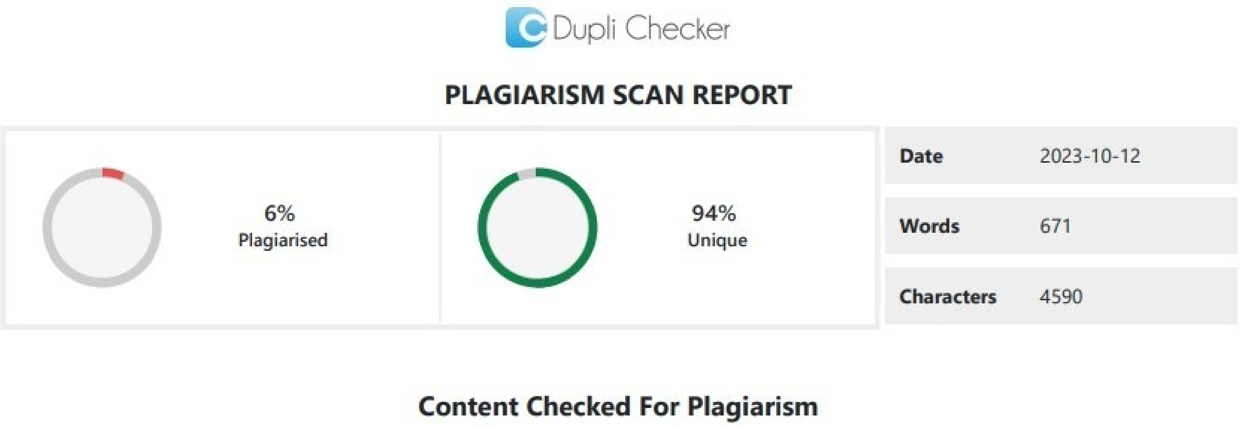
6.1 Feature Scope:

We all know that any Analysis is not perfect and it consists of some limitations which define the Future scope of the Research Work. This project work also contains some limitations which we are considering as the Future Scope of the Project. We can also describe the data in other formats like Geographical format where we can depict the countries on the World map. We can also apply various Machine Learning Algorithms to the data set after Analysis and can create a Predictive Model which can predict the statistics of the Future Olympic Games.

Chapter 7

Plagiarism Report

Presenting work or ideas from another source as your own, with or without consent of the original author, by incorporating it into your work without full acknowledgement.



Dupli Checker

PLAGIARISM SCAN REPORT

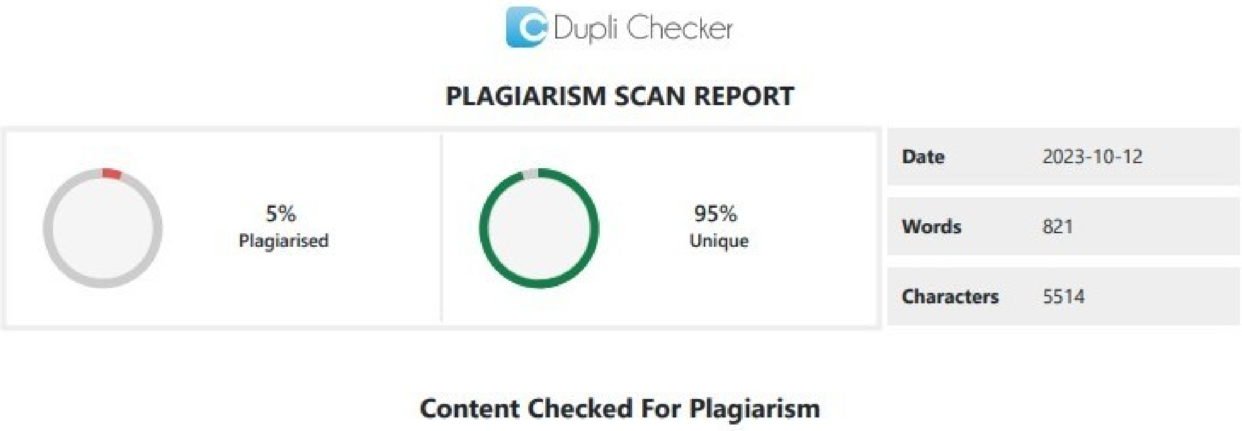
6% Plagiarised

94% Unique

Date	2023-10-12
Words	671
Characters	4590

Content Checked For Plagiarism

Figure 7.1: Plagiarism Report for Literature Survey



Dupli Checker

PLAGIARISM SCAN REPORT

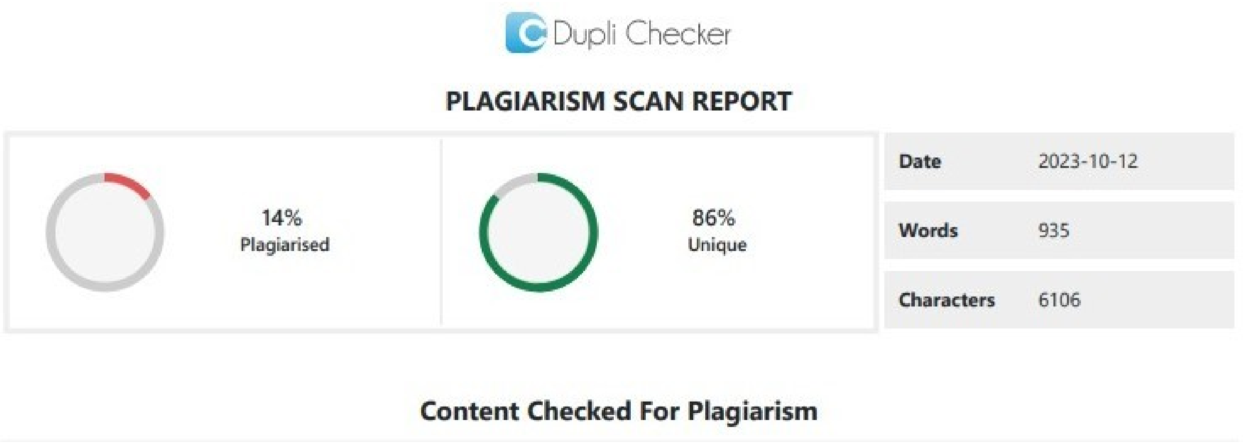
5% Plagiarised

95% Unique

Date	2023-10-12
Words	821
Characters	5514

Content Checked For Plagiarism

Figure 7.2: Plagiarism Report for Literature Survey.



Dupli Checker

PLAGIARISM SCAN REPORT

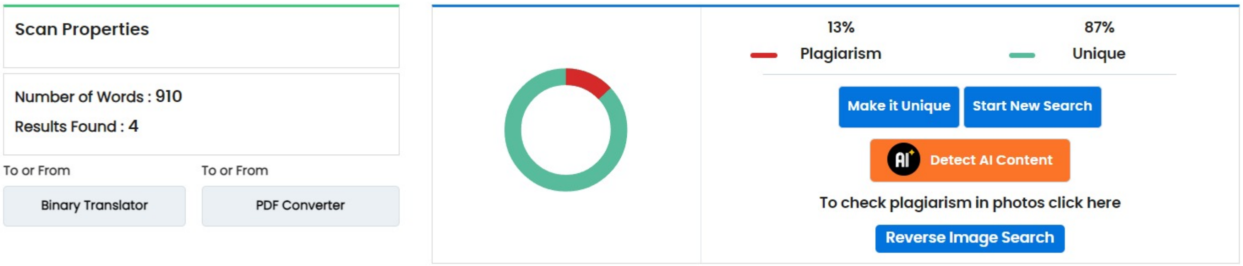
14% Plagiarised

86% Unique

Date	2023-10-12
Words	935
Characters	6106

Content Checked For Plagiarism

Figure 7.3: Plagiarism Report for Research Paper



Scan Properties

Number of Words : 910

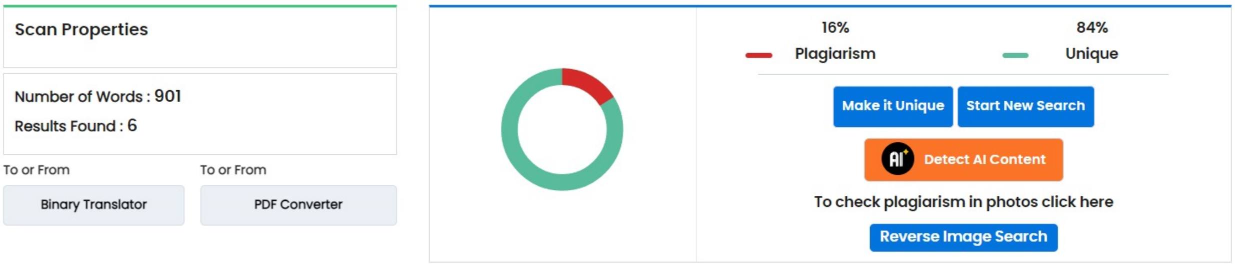
Results Found : 4

13% Plagiarism

87% Unique

To check plagiarism in photos click here

Figure 7.4: Plagiarism Report for Research Paper



Scan Properties

Number of Words : 901

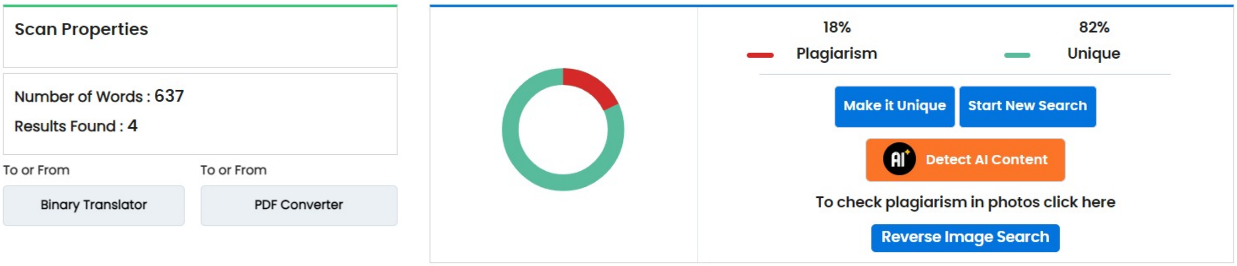
Results Found : 6

16% Plagiarism

84% Unique

To check plagiarism in photos click here

Figure 7.5: Plagiarism Report for Content



Scan Properties

Number of Words : 637

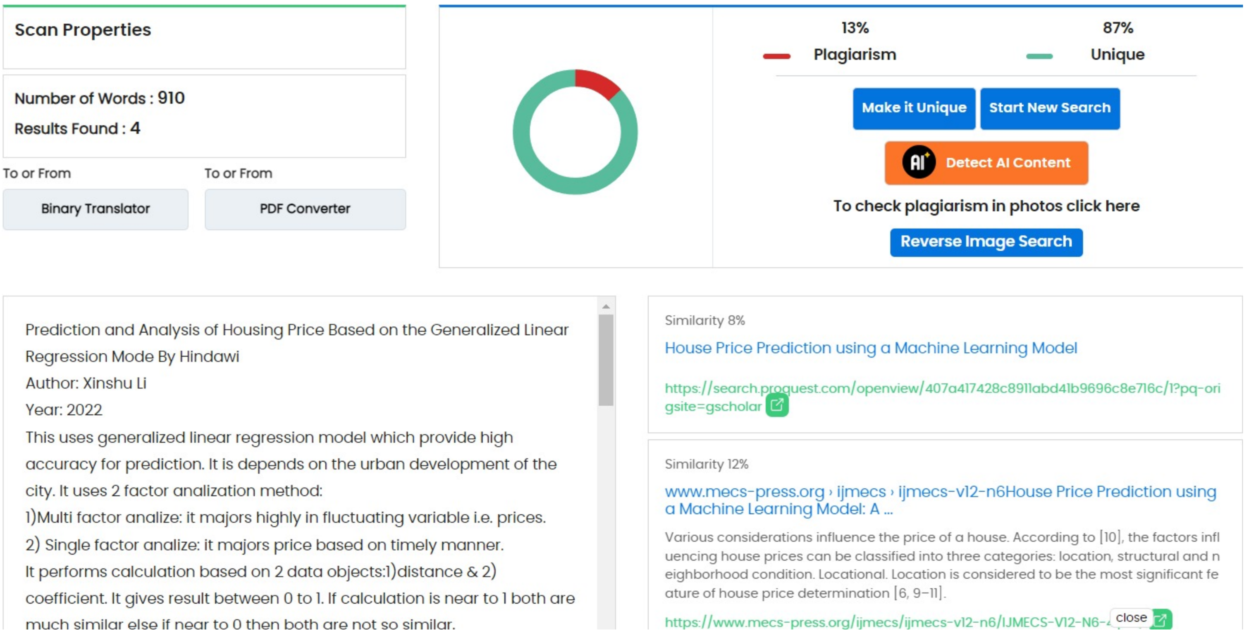
Results Found : 4

18% Plagiarism

82% Unique

To check plagiarism in photos click here

Figure 7.6: Plagiarism Report for Content



Number of Words : 910

Results Found : 4

To check plagiarism in photos click here

Prediction and Analysis of Housing Price Based on the Generalized Linear Regression Mode By Hindawi
Author: Xinshu Li
Year: 2022
This uses generalized linear regression model which provide high accuracy for prediction. It is depends on the urban development of the city. It uses 2 factor analization method:
1)Multi factor analize: it majors highly in fluctuating variable i.e. prices.
2) Single factor analize: it majors price based on timely manner.
It performs calculation based on 2 data objects:1)distance & 2) coefficient. It gives result between 0 to 1. If calculation is near to 1 both are much similar else if near to 0 then both are not so similar.

Similarity 8%

House Price Prediction using a Machine Learning Model

https://search.proquest.com/openview/407a417428c8911abd41b9696c8e716c/1?pq-origsite=gscholar

Similarity 12%

www.mecs-press.org › ijmecs › ijmecs-v12-n6House Price Prediction using a Machine Learning Model: A ...

Various considerations influence the price of a house. According to [10], the factors influencing house prices can be classified into three categories: location, structural and neighborhood condition. Locational. Location is considered to be the most significant feature of house price determination [6, 9–11].

https://www.mecs-press.org/ijmecs/ijmecs-v12-n6/IJMECS-V12-N6-

close

Figure 7.7: Plagiarism Report for Research Paper.

References

1. Wikipedia contributors: https://en.m.wikipedia.org/wiki/Olympic Games, last accessed 2020/11/02.

2. Dey S K, Rahman M M, Siddiqi U R and Howlader A 2020 Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach J. Med. Virol. 92 632–8

3. Cutait, M.: Management performance of the Rio 2016 Summer Olympic Games. Research Paper submitted and approved to obtain the Master's degree in Sports Administration at AISTS in Lausanne, Switzerland

4. Bondu R, Cloutier V, Rosa E and Roy M 2020 An exploratory data analysis approach for assessing the sources and distribution of naturally occurring contaminants (F, Ba, Mn, As) in groundwater from southern Quebec (Canada) Appl. Geochem. 114 104500

5. Moreno A, Moragas M and Paningua R 1999 The evolution of volunteers at the Olympic Games Proceedings of Symposium on Volunteers (Lausanne, Switzerland: Global Society and the Olympic Movement) pp 1–18

6. Abeza G, Braunstein-Minkove J R, S ́eguin B, O'Reilly N, Kim A and Abdourazakou Y 2020 Ambush marketing via social media: The case of the three most recent Olympic Games Int. J. Sport Communication 1–25

7. Yamunathangam D, Kirthicka G and Shahanas P 2018 Performance Analysis in Olympic Games using Exploratory Data Analysis Techniques International Journal of Recent Technology and Engineering (IJRTE) 7 251–3

8. Wikipedia contributors: Exploratory data analysis, https://en.wikipedia.org/wiki/Exploratory data analysis, last accessed 2020/11/11

9. Ramachandran K. M. and Tsokos C P 2020 Mathematical statistics with applications in R (Academic Press)

10. Lange D Summer Olympics: number of participating countries 1896-2016 Statista.com