Technical Report
for

Data Journey: SQL, Anomaly Detection, and Power BI

Written by:

Mays Moh’d Salih Al-Fasfous

October 2023

Data Analytics & SQL Server Training

Abstract

In an increasingly data-centric world, the role of data analytics is pivotal for informed decision-making, streamlined processes, and enhancing user experiences. This technical report presents a comprehensive overview of three distinct data analytics projects, each showcasing specialized skills in database design, predictive analytics, and data visualization using SQL Server, Python, and Microsoft Power BI. These projects serve as exemplars of how data-driven insights can empower organizations and individuals across diverse domains.
The first project centers on the creation of an Online Ticket Booking Database System, where meticulous database design and implementation were prioritized to ensure data consistency, integrity, and security. The primary objective was to simplify the online ticket booking process for users while upholding the highest standards of data quality. This endeavor not only enhanced the user experience but also underscored the critical role a well-structured database plays in optimizing operational efficiency.

The second project delves into descriptive and predictive analytics, utilizing the IoTID20 dataset to categorize anomalies. Through data preparation, visualization, and the application of predictive models, this project highlighted the importance of selecting appropriate analytical techniques and feature selection methods. The findings demonstrated how data analytics can provide actionable insights and inform data-driven decision-making, particularly in the context of anomaly detection.

The third project leveraged the power of Microsoft Power BI to create impactful Olympic data dashboards. These dashboards addressed Women's Olympic Participation, Jordan's Performance, and Seasonal Olympic Insights, offering valuable information from a rich dataset. Notably, they emphasized gender equality, performance enhancement, and strategic recommendations for Winter Olympic hosting countries. This project exemplified the capacity of data visualization to inform decision-makers and improve the Olympic experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

In today's data-driven world, data analytics is essential for informed decision-making and user-friendly insights. Whether building databases or working with existing data, data integrity and consistency are paramount for accurate analytics. As a data analytics specialist, I've completed three projects showcasing my skills in data analytics and SQL Server This report will present three distinct projects, each highlighting specific aspects of data analytics, database design, and implementation using SQL Server, as well as the creation of Power BI dashboards.

2 Online Ticket Booking Database System

2.1 Background and Objectives

The Online Booking System project centers on the design and implementation of a robust relational database management system (RDBMS). The primary objective is to design and develop a robust database capable of efficiently handling user bookings for movie shows and simplifying the online ticket booking process. This endeavor aimed to ensure data consistency, integrity, and security while enhancing the overall user experience.

2.2 Database Design

2.2.1 Scenario

A comprehensive Online Ticket Booking System is designed to simplify event ticket reservations. Customers can create accounts, search for shows and events, select their preferred seats, and make secure bookings. Each event, referred to as a "Show," is defined by its date, time, movie details, and theater. Administrators oversee the system, manage customer accounts, and monitor bookings. Bookings include customer details, ticket information, and payment status, ensuring a seamless experience for all users.

2.2.2 Conceptual Model

In the conceptual model, I have identified four key entities: Ticket, Customer, Admin, and Show.

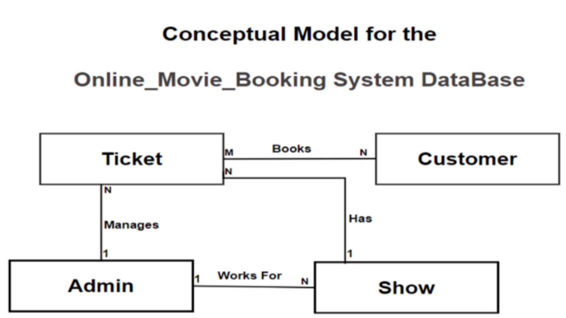

Figure 2: Conceptual Model for Online Movie Booking System Database

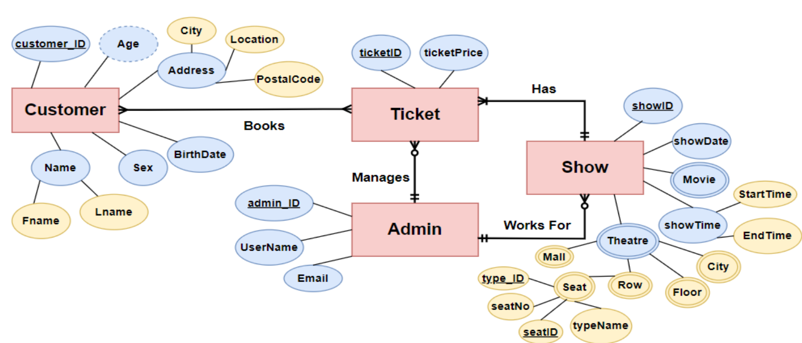

Figure 2: Conceptual Model with Attributes for Online Movie Booking System Database

2.2.3 Schema, Mapping and Normalization

In Figure 2, you can see derived, composite, and multivalued attributes, as well as many-to-many (M:N) relationships between entities. After applying mapping rules and achieving the 1st Normal Form (1st NF), I identified partial dependencies in "Seats," which prompted me to apply the 2nd Normal Form (2nd NF). The final schema result is shown in Figure 3.

Customers (customer_ID, Name, BirthDate, Email, Password, MobileNo, Sex, City, Location, PostalCode)
Tickets (ticket_ID, Ticket_Num, Price, show_StartTime, show_EndTime, show_Date, show_ID, theatre_ID)
Shows (show_ID, show_StartTime, show_EndTime, show_Date, theatre_ID, Language, movie_ID)
Movies (movie_ID, movie_Name, movie_Description, movie_Age, movie_Duration, movie_Genre)
Theatres (theatre_ID, theatre_Name, city_ID, floor_ID, mall_ID)
Cities (city_ID, city_Name)
Malls (mall_ID, mall_Name)
Floors (floor_ID, floor_Name)
Rows_Theatre (row_ID, row_Name, thetre_ID)
Seats (seat_ID, seat_No, row_ID , seatType_ID)
Seat_Types (seatType_ID, seatType_Name)
Admins (admin_ID, Email, Password, Username, Sex)
Bookings (booking_ID, customer_ID, ticket_ID, booking_Date, has_paid_ticket, quantity, admin_ID)

Note: Primary keys are underlined, and foreign keys are highlighted in gray.

Figure 3: Displaying a relational database schema and its constraints

2.2.4 Logical and Physical Model

After finalizing the schema and ensuring integrity and consistency through mappings and normalization, I created a logical model (Figure 4), which represents primary key (P.K) and foreign key (F.K) constraints and attributes, and then I defined attribute constraints and data types in the physical model (Figure 5), as shown below:

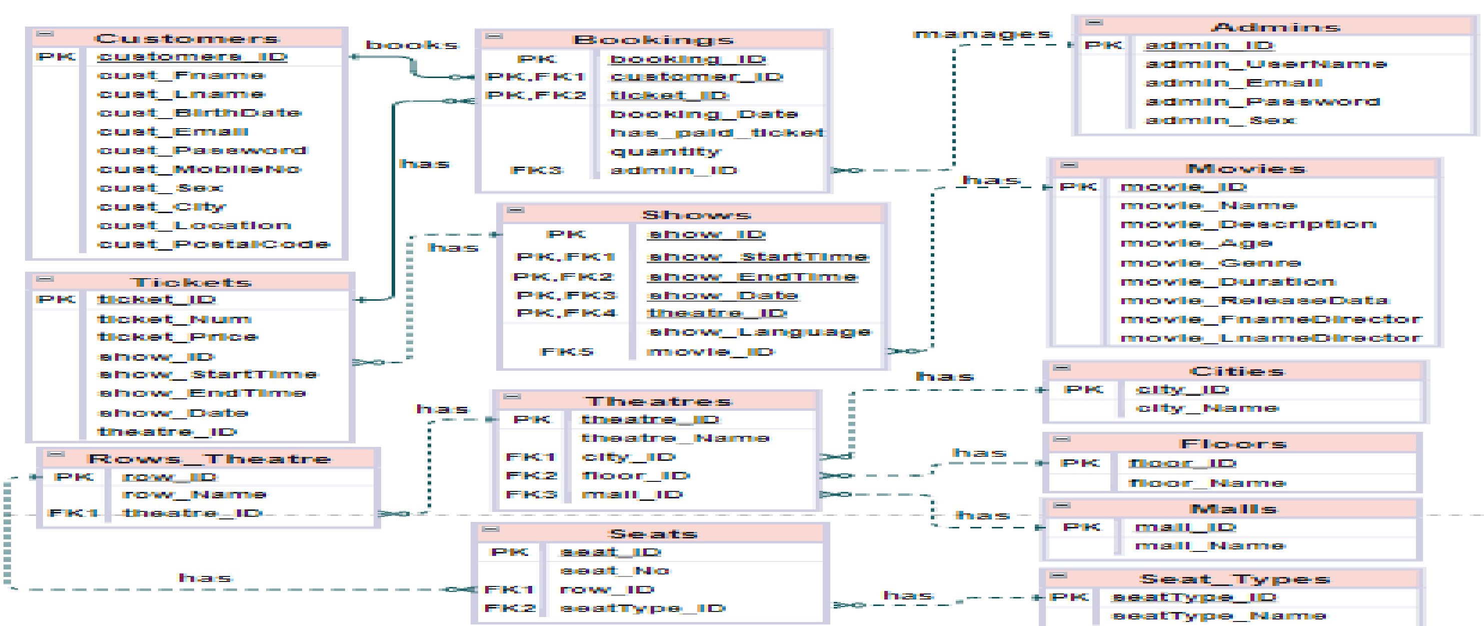

Figure 4: Logical Model for Online Movie Booking System Database

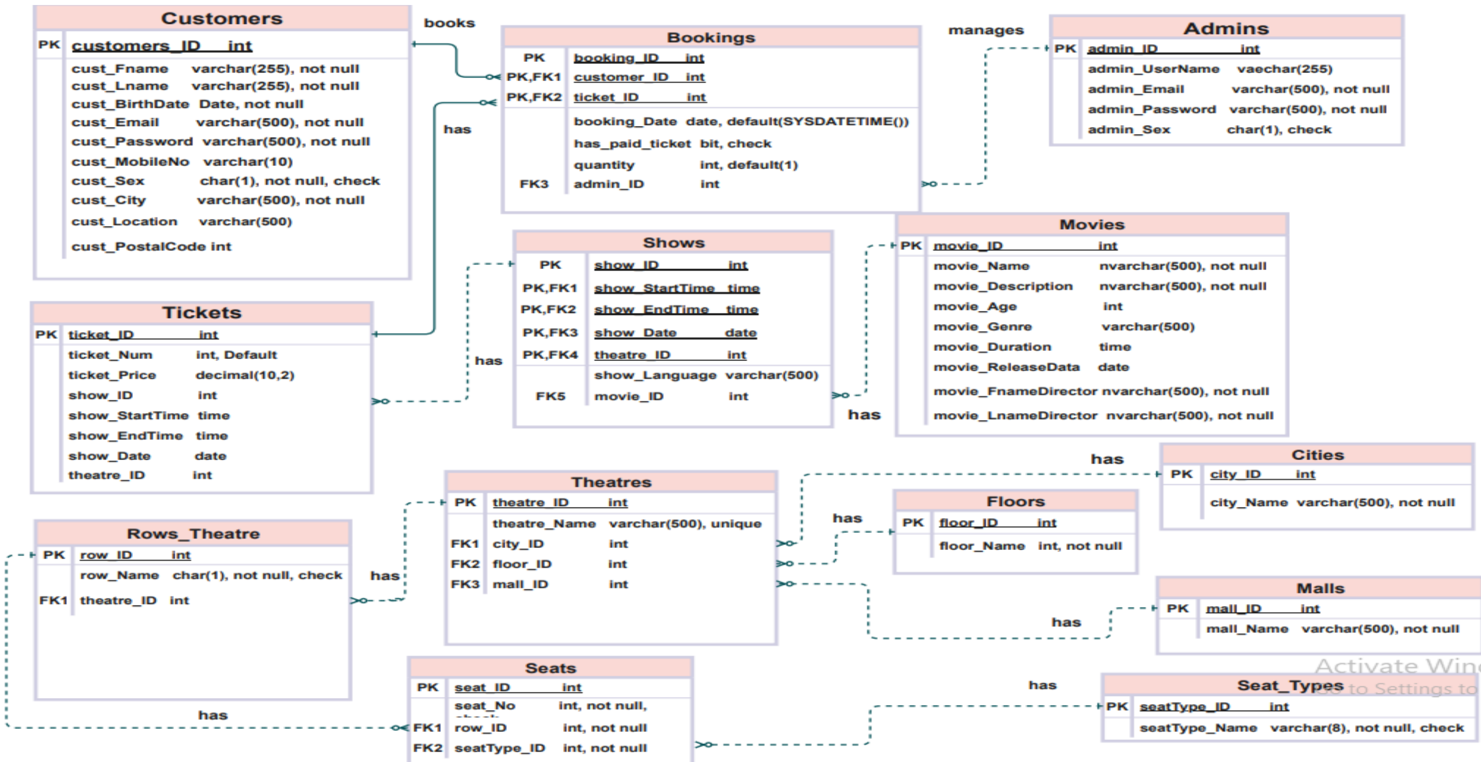

Figure 5: Physical Model for Online Movie Booking System Database

2.3 Database Implementation

In this phase, we translated our logical and physical models into a functional database using Microsoft SQL Server as the chosen Database Management System (DBMS). This involved defining the database schema, specifying tables, relationships, and key SQL functionalities:

- **Data Retrieval**: I employed SELECT statements, JOINs, WHERE clauses, and single row functions for efficient data retrieval from multiple tables, to filter results based on specified conditions, and to customize output.
- **User Management**: SQL Server's CREATE USER statements enabled secure user access.
- **Stored Procedures**: I created procedure to facilitate data management operations.
- **Functions**: Custom functions were created for reusable logic.
- **Data Manipulation**: INSERT, UPDATE, and DELETE statements facilitated data management.
- **Security**: SQL Server features ensured user account security and permission control.

2.4 Results and Outcomes

2.4.1 Data Integrity and Normalization

- Schema Design and Constraints: The database schema was meticulously designed to enforce data accuracy and reliability. Constraints were applied to maintain data integrity.
- Normalization: The schema was normalized up to the 2nd Normal Form (2nd NF), effectively managing complex relationships and minimizing data redundancy.

2.4.2 Alignment with User and System Requirements

- User Requirements: The system enables customers to effortlessly book tickets, select preferred seats, and enjoy a streamlined experience. Administrators efficiently manage bookings.
- System Requirements: The system meets scalability and security requirements.

2.4.3 SQL Queries and RDBMS

- SQL Queries: Queries, including JOINs, retrieve data efficiently.
- Microsoft SQL Server: SQL Server provides a stable and secure platform for all database operations.

2.4.4 Data Extraction and Tools

- Query tools facilitate the extraction of meaningful data for analysis, further enhancing the system's utility.

2.5 Conclusion

In conclusion, the Online Ticket Booking Database System has successfully achieved its primary objective of providing a robust and efficient platform for handling user bookings while simplifying the online ticket booking process. It aligns seamlessly with user and system requirements, upholds data integrity, and leverages the power of SQL queries and Microsoft SQL Server for efficient operations. It stands as compelling evidence of the critical role a well-structured database plays in enhancing user experiences and optimizing operational efficiency.

3 Descriptive and Predictive Analytics for Anomaly Categories

3.1 Background and Objectives

This project focuses on analyzing the IoTID20 Dataset using descriptive and predictive analytics techniques within the Python programming language. Our primary focus is on the 'Cat' column, which categorizes anomalies. The specific objectives include data preparation, exploration of data relationships through visualization, application of predictive models, and the comparative assessment of various predictive analytical techniques to achieve accurate anomaly categories prediction. I aim to demonstrate how descriptive analytics can provide actionable insights specific to this dataset and underline the critical role of selecting the most suitable predictive techniques for anomaly categories prediction. Additionally, I explore how these analytics enhance data-driven decision-making for anomaly categories detection.

3.2 Data Source

The dataset used for this project is sourced from the IoTID20 dataset (IoT Intrusion Dataset). This dataset was specifically designed for detecting anomalous activity in IoT networks, featuring a comprehensive set of network and flow-based features. IoTID20 serves as a valuable resource for precise anomaly detection within IoT networks.

3.3 Data Preparation

In the initial phase, we loaded the dataset into a pandas data frame, followed by cleaning, where we removed irrelevant features, handled duplicates, and ensured data quality. The 'info()' function identified data types, revealing that all columns, except 'Cat', were of float64 or int64 types. 'Cat' contains categorical data and is classified as an object type. This meticulous data preparation readied the dataset for subsequent analysis, ensuring its relevance and integrity.

3.4 Descriptive Analytics

3.4.1 Descriptive Analytic Techniques

In this solo project, I conducted a detailed analysis of the 'Flow_Duration' and 'Idle_Mean' features using descriptive analytic techniques. The analysis included computations of essential statistics and utilization of density charts for data visualization and outlier detection. Additionally, we explored the 'Cat' column, representing network intrusion categories, using measures of frequency (absolute frequency) and measures of central tendency (mode). The table and chart below display these results, shedding light on data characteristics. These insights are instrumental in understanding the potential impact of these features on anomaly categorization in our project.

Table 1: Descriptive Analytic Techniques for numerical columns

	Measures	Flow_Duration	Idle_Mean
0	Central Tendency (Mean)	1310.317983	884.064876
1	Central Tendency (Median)	149.000000	83.000000
2	Dispersion (Range)	99984.000000	99973.000000
3	Dispersion (Variance)	42546457.499465	13367798.360496
4	Dispersion (Standard Deviation)	6522.764560	3656.199989
5	Dispersion (Interquartile Range)	220.000000	117.000000
6	Dispersion (Coefficient of Variation)	4.977986	4.135657
7	Position (Quartile (Q1))	78.000000	34.000000
8	Position (Quartile (Q3))	298.000000	151.000000
9	Position (Quartile (Q2))	149.000000	83.000000

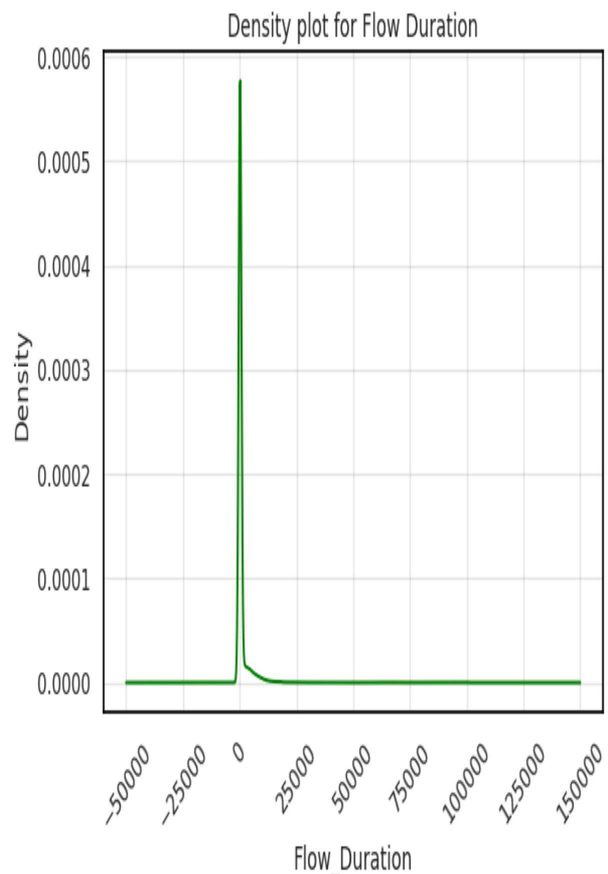

Figure 6: Density plot for Flow_Duration

The most frequent 'Cat' column value is 'Mirai', as determined by Mode, and this is visually represented in the following charts that display Absolute Frequency for 'Cat' column :

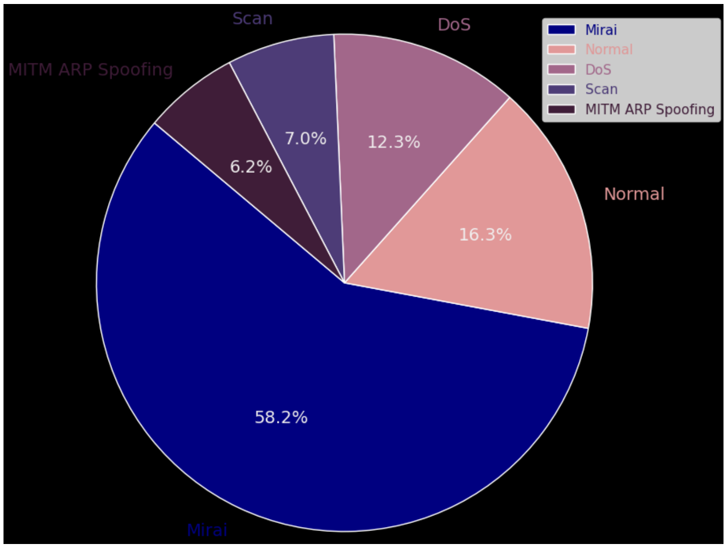

Figure 7: Distribution of Cat Column using Pie Chart

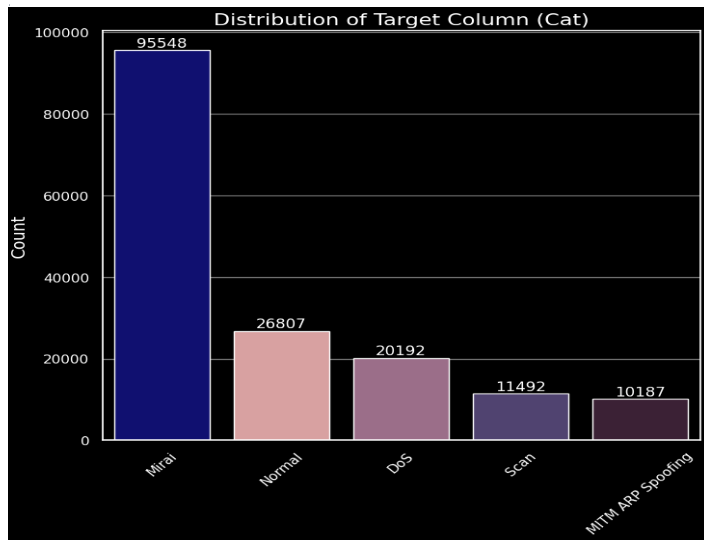

Figure 8: Distribution of Cat Column using Bar Chart

3.4.2 Insights - Visualizations using Python

In this section, I explored insights derived from visualizations using Python, focusing on the 'Cat' column and its relationships with other features. A bar chart of the average (mean) 'Flow Duration' shows that the DoS category in the 'Cat' column has the highest values. I also displayed a pair plot illustrating the relationships between 'Flow_Duration', 'Active_Mean', and 'Idle_Mean' when 'Cat' is DoS. These visuals offer insights into how these features affect anomaly categorization.

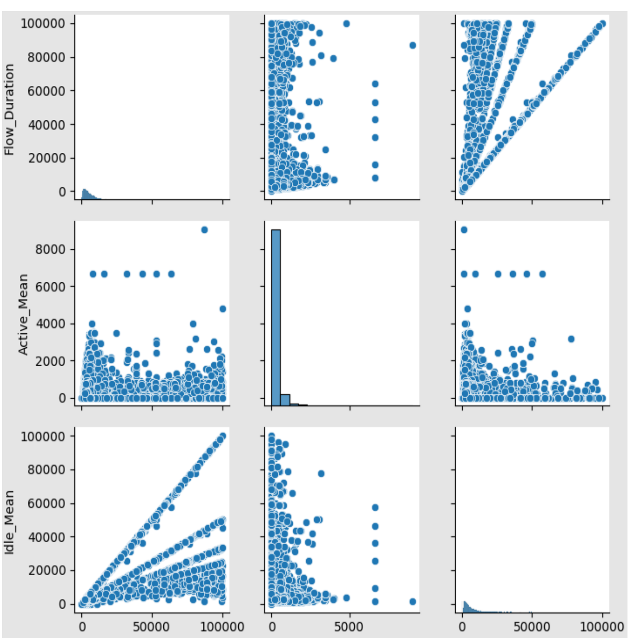

Figure 9: Pair plot Relationships between the Features for Category DoS

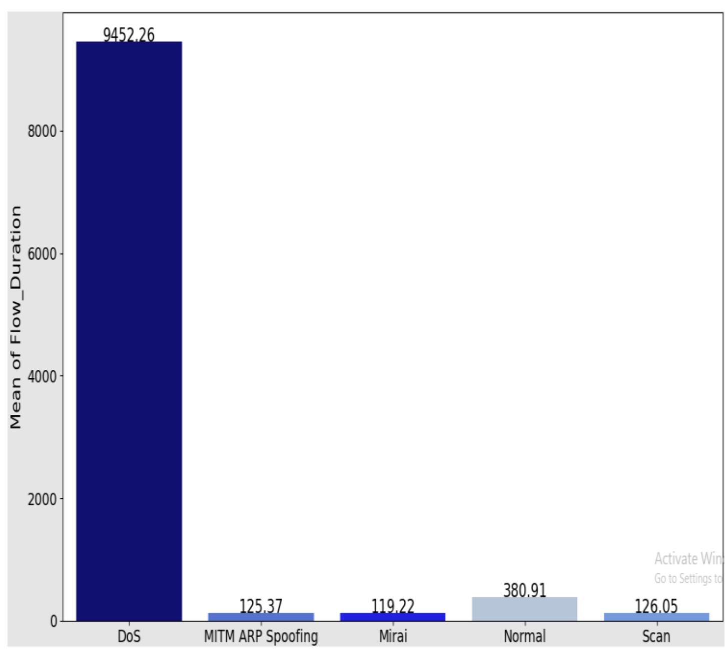

Figure10: Bar plot for Mean of Flow_Duration by Category

In addition to the insights gained from the 'DoS' category, I also examined the Mirai category within the 'Cat' column. A bar chart of the average (mean) 'Fwd_Pkts/s' revealed that the Mirai category exhibited the highest values, as shown in the forthcoming chart. Additionally, a pair plot demonstrated the relationships between 'Fwd_Pkts/s,' 'Subflow_Fwd_Pkts,' and 'Fwd_Act_Data_Pkts' when 'Cat' is 'Mirai.' These visualizations provide a deeper understanding of how these features impact anomaly categorization.

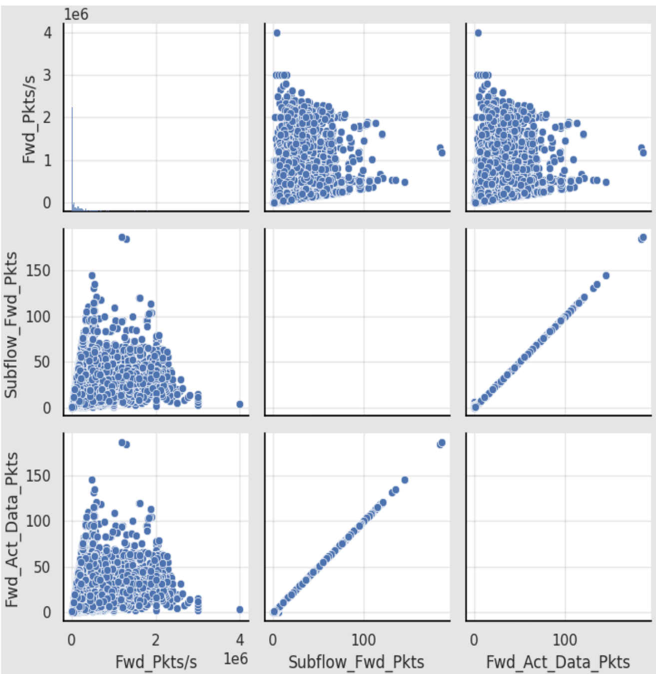

Figure 11: Pair plot for Relationships between the Features for Category Mirai

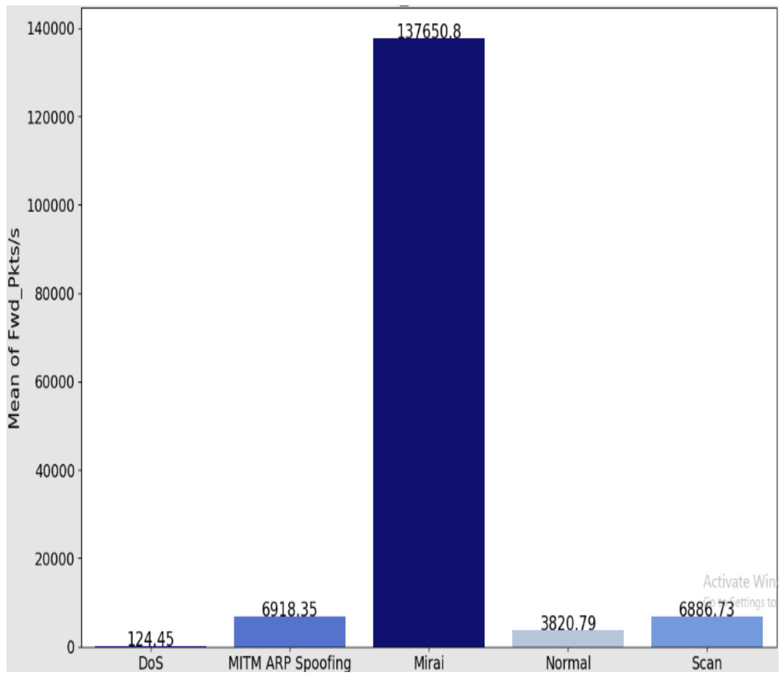

Figure 12: Bar plot for Mean of Fwd_Pkts/s by Category

3.5 Predictive Analytics

3.5.1 Classification Algorithms

In this solo project, various classification algorithms were applied using Python and the Scikit-learn library to predict anomaly categories within the 'Cat' column accurately. A detailed performance comparison table showcases evaluation measures for these models, clearly indicating that K-Nearest Neighbours (KNN) is the best-performing model for the anomaly prediction and categorization task.

Table 2: Comparison of Metrics for Different Algorithms

	Models	Accuracy	Precision	Recall	F1-Score
0	SVC	0.840546	0.788565	0.840546	0.805021
1	KNN	0.885199	0.874795	0.885199	0.877721
2	Decision Tree	0.868339	0.870662	0.868339	0.869042
3	GaussianNB	0.620651	0.858077	0.620651	0.674882
4	Logistic Regression	0.849129	0.788453	0.849129	0.805289
5	Random Forest	0.869718	0.865322	0.869718	0.867387

3.5.2 Feature Selection

In this project, I enhanced the performance of K-Nearest Neighbours (KNN) and Decision Tree Classifier (DT) using feature selection techniques, resulting in improved accuracy for both models. The comparison tables illustrate the results, demonstrating that all algorithms performed better with feature selection. DT stood out as the top performer, achieving an impressive 90% accuracy and recall rate. This underscores the effectiveness of feature selection in improving model performance, with DT emerging as the preferred choice for anomaly prediction and categorization.

Table 3: Comparison for Different Feature Selection Techniques – KNN

	Feature Selection	Accuracy	Precision	Recall	F1-Score
0	VarianceThreshold	0.885403	0.875026	0.885403	0.877920
1	SelectKBest	0.883615	0.869232	0.883615	0.873600
2	Select Percentile	0.835181	0.784363	0.835181	0.795258
3	Generic Univariate Select	0.883615	0.869232	0.883615	0.873600
4	SelectFwe	0.885199	0.874610	0.885199	0.877594
5	SelectFpr	0.885148	0.874610	0.885148	0.877572
6	SelectFdr	0.885148	0.874610	0.885148	0.877572

Table 4: Comparison for Different Feature Selection Techniques – DT

	Feature Selection	Accuracy	Precision	Recall	F1-Score
0	VarianceThreshold	0.902161	0.894511	0.902161	0.896405
1	SelectKBest	0.872069	0.871668	0.872069	0.871649
2	SelectFromModel	0.871353	0.870211	0.871353	0.870447
3	Recursive feature elimination	0.901804	0.893947	0.901804	0.895904
4	Select Percentile	0.841465	0.784238	0.841465	0.794019
5	Generic Univariate Select	0.895417	0.885304	0.895417	0.888284
6	SelectFwe	0.902417	0.894703	0.902417	0.896636
7	SelectFpr	0.902263	0.894598	0.902263	0.896497
8	SelectFdr	0.901855	0.894161	0.901855	0.896160

Table 5: Model Performance with and without Feature Selection - KNN

	K-Nearest Neighbours (KNN)	Accuracy	Precision	Recall	F1-Score
0	KNN With VarianceThreshold	0.885403	0.875026	0.885403	0.877920
1	KNN Without VarianceThreshold	0.885199	0.874795	0.885199	0.877721

Table 6: Model Performance with and without Feature Selection - DT

	Decision Tree Classification (DT)	Accuracy	Precision	Recall	F1-Score
0	DT with SelectFwe	0.902417	0.894703	0.902417	0.896636
1	DT Without SelectFwe	0.868646	0.871274	0.868646	0.869474

Table 7: Comparison of KNN and DT Performance

	Models	Accuracy	Precision	Recall	F1-Score
0	DT With With SelectFwe	0.902417	0.894703	0.902417	0.896636
1	KNN With VarianceThreshold	0.885403	0.875026	0.885403	0.877920

3.6 Results and Outcomes

- In the context of the 'DoS' category, I observed through the descriptive analytics that the 'Flow_Duration' feature plays a significant role, indicating its importance in anomaly categorization for this specific category. As a result, a decision can be made to prioritize 'Flow_Duration' in the feature selection process, considering it as a key attribute for enhancing the predictive models.
- For the 'Mirai' category, the descriptive analytics highlighted 'Fwd_Pkts/s' as a prominent feature, suggesting its relevance in anomaly detection. Consequently, it can include 'Fwd_Pkts/s' in the feature selection, recognizing its potential to improve model accuracy for 'Mirai' category classification.
- The application of feature selection techniques significantly enhanced the performance of both the K-Nearest Neighbours (KNN) and Decision Tree Classifier (DT) models.
- The outstanding performance of the Decision Tree Classifier (DT) models with the SelectFwe feature selection, achieving an accuracy and recall rate of 90%, led us to designate it as the top-performing model overall, especially for anomaly prediction and categorization tasks.
- These findings underscore the importance of feature selection in improving model accuracy and emphasize DT as the optimal choice for my project's objectives.

3.7 Conclusion

In this project, we conducted a comprehensive analysis of the IoTID20 Dataset using descriptive and predictive analytics in Python. Our findings underscore the importance of 'Flow_Duration' for detecting 'DoS' anomalies and 'Fwd_Pkts/s' for identifying 'Mirai' anomalies. While K-Nearest Neighbours initially excelled as a predictive model, the inclusion of feature selection elevated the Decision Tree Classifier to a remarkable 90% accuracy and recall rate. This project illustrates the efficacy of analytics in anomaly category detection, with a focus on the vital role of feature selection and the considerable potential of the Decision Tree Classifier.

4 Exploring Olympic Data: Visualizations and Insights with Power BI

4.1 Background and Objectives

In this solo project, I used Microsoft Power BI to create three meaningful Olympic data dashboards. These dashboards explore Women's Olympic Participation, Jordan's Performance, and Seasonal Olympic Insights, making data more engaging. I've designed these dashboards to be clear and insightful, offering valuable information from a large dataset. This report highlights each dashboard's insights and importance in decision-making.

4.2 Data Source

This project utilizes the Olympic dataset, a compilation of three tables: Sport_Details, noc_region, and dataset_olympics, to explore Olympic history and trends.

4.3 Data Preparation

The Olympic dataset underwent essential data preparation in Power BI, including handling missing values, adjusting data types, using the first row as headers, validating data quality, and adding custom columns. These steps ensured dataset accuracy for Power BI dashboard visualization.

4.4 Insights - Visualizations using Power BI

4.4.1 Women's Participation in Olympic Dashboard

This Power BI dashboard project consists of two pages, each telling a distinct story. The first page, titled "Equality in Olympic Participation," focuses on gender participation equality and provides insights into the distribution of male and female participants. The second page, "Women's Olympic Events Analysis and Regional Insight," delves into women's participation in Olympic events and offers regional perspectives. The audience for this project is the Olympic organization, with a particular interest in addressing gender participation disparities and enhancing the Olympic experience for women.

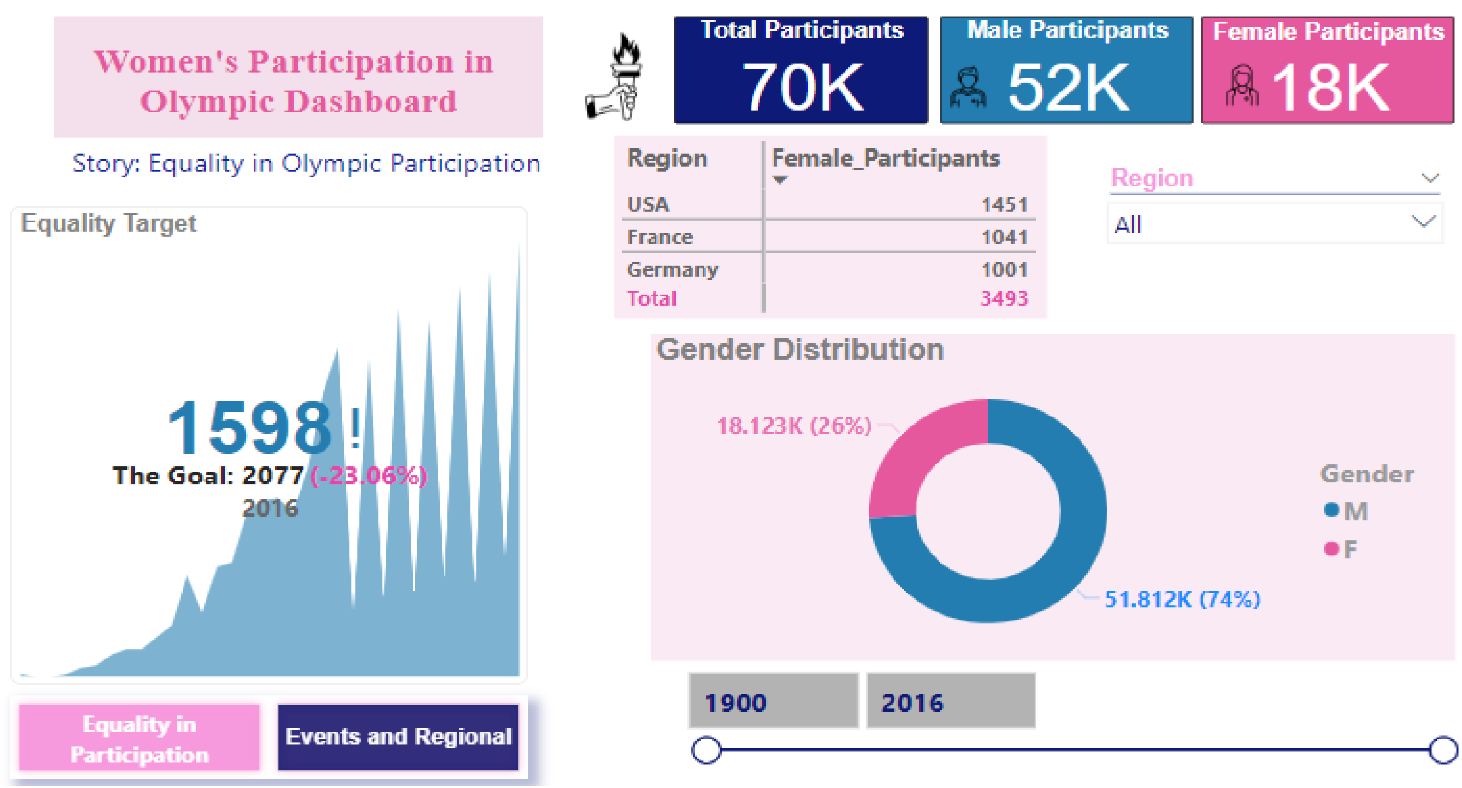

Figure 13: Women's Participation in Olympic Dashboard – Equality in Participation page

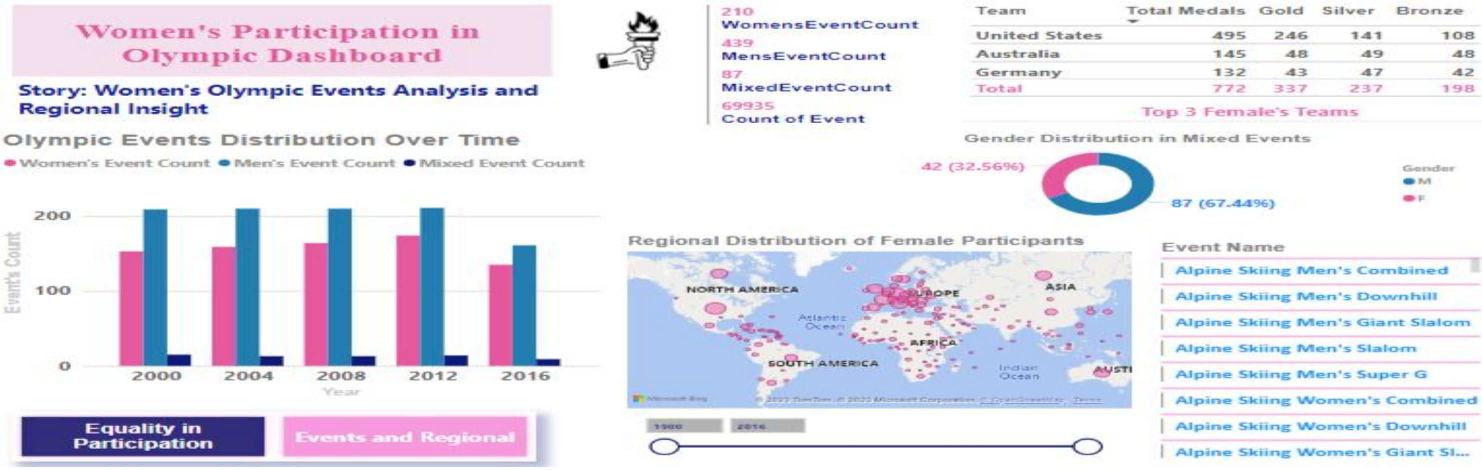

Figure 14: Women's Participation in Olympic Dashboard-Events and Regioal page

4.4.1.1 First Page – Story: Equality in Olympic Participation

- **Total participants:** The first page reveals that, in general, male participants outnumber female participants, indicating gender participation disparities.
- **Target KPI:** The target of achieving equal participation between men and women, while set, has not yet been reached as of 2016.
- **Top 3 countries:** The dashboard highlights the top three countries with the highest total female participants: USA, France, and Germany.
- Insights: By utilizing slicers for region and year, the organization can assess progress toward gender participation equality in specific regions and over different timeframes.

4.4.1.2 Second Page – Story: Women's Olympic Events Analysis and Regional Insight

- Event counts: The second page provides insights into the distribution of Olympic events, indicating that men's events significantly outnumber women's and mixed events.
- Event distribution over time: A clustered column chart visually represents the distribution of Olympic events over time, highlighting the need to increase women's events.
- Regional distribution: A map chart reveals regional distributions of female participants, with the USA having the highest number of female participants.
- Gender distribution in mixed events: A donut chart clearly shows that male participants outnumber female participants in mixed events.
- Insights: The insights from this page suggest a preference for female participants to compete in women's events rather than mixed events. The organization can leverage this information for marketing and engagement strategies.

4.4.2 Jordan's Olympic Performance Analysis Dashboard

This section presents the results and outcomes derived from the analysis of Jordan's Olympic Performance, focusing on the dashboard titled "Analyzing Jordan's Olympic Journey". This analysis is aimed at providing actionable insights to the Jordan Olympic Committee to enhance the country's performance in the Olympic Games.

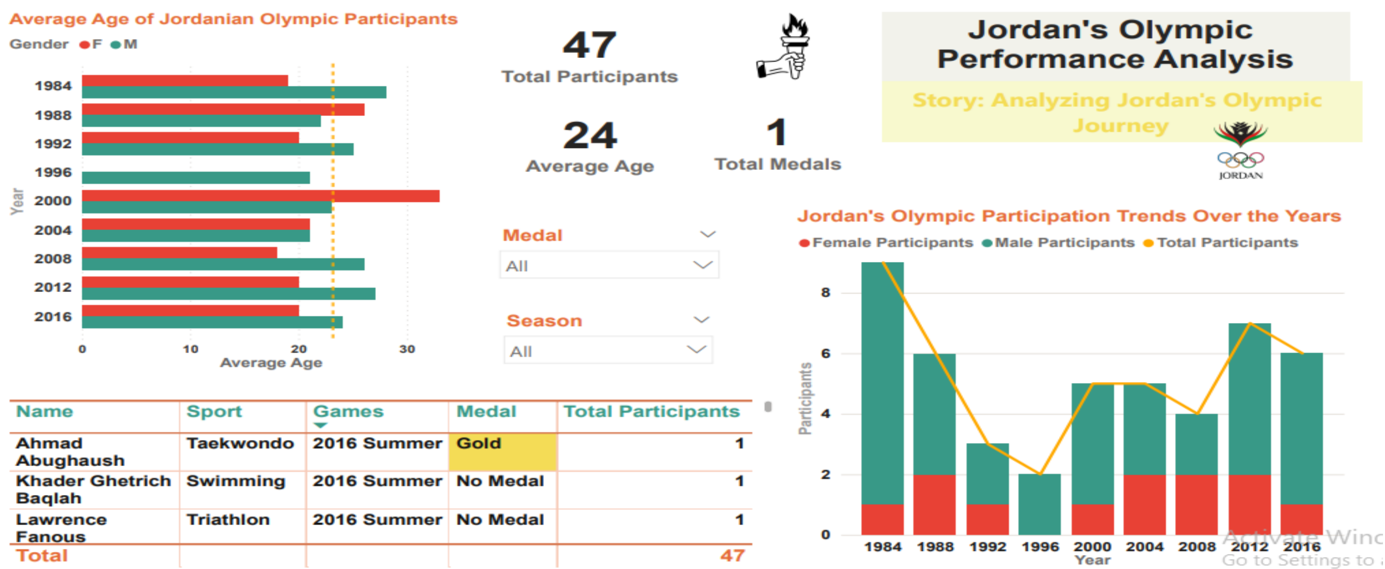

Figure 15: Jordan's Olympic Performance Analysis Dashboard

4.4.2.1 Story: Analyzing Jordan's Olympic Journey

- Total Participants: From 1984 to 2016, Jordan has participated in the Olympics with a total of 47 athletes. However, there is significant potential for growth in the number of participants.
- Medals: Jordan has won a total of 1 gold medal in the Taekwondo sport during the 2016 Summer Games, represented by Ahmad Abughaush.
- Average Age: The average age of Jordanian Olympic participants, both male and female, was approximately 24 years. Users can interactively select specific years and genders to view corresponding data.
- Seasons: The analysis reveals that Jordan has exclusively participated in the Summer Olympics, with no presence in the Winter Games.
- Single Event Participation: The data indicates that most Jordanian participants have competed in only one event during each Olympic Games.

4.4.3 Seasonal Olympic Insight Dashboard

This section presents the results and outcomes derived from the analysis of the "Winter Olympic Hosts" dashboard, designed for countries hosting the Winter Olympics. The objective is to provide insights and recommendations for countries hosting the Winter Games.

4.4.3.1 Story: Winter Olympic Hosts

- Seasonal Sport Distribution: A donut chart highlights the distribution of sports by season, revealing that Summer Olympics have a significantly higher percentage (54%) of sports compared to Winter Olympics (18%).
- Top Winter Hosts: A table lists the top three Winter Olympic hosts by the total number of participants, with the USA, Canada, and France leading the rankings.
- Gauge for Gold Medals: The gauge chart sets a target for the number of Winter gold medals to be equal to that of summer gold medals.
- Season and Event Slicer: Users can select specific seasons and events using slicers to compare event counts, average durations and the total number of Olympic participants, distinguishing between those in the Winter and Summer Games.

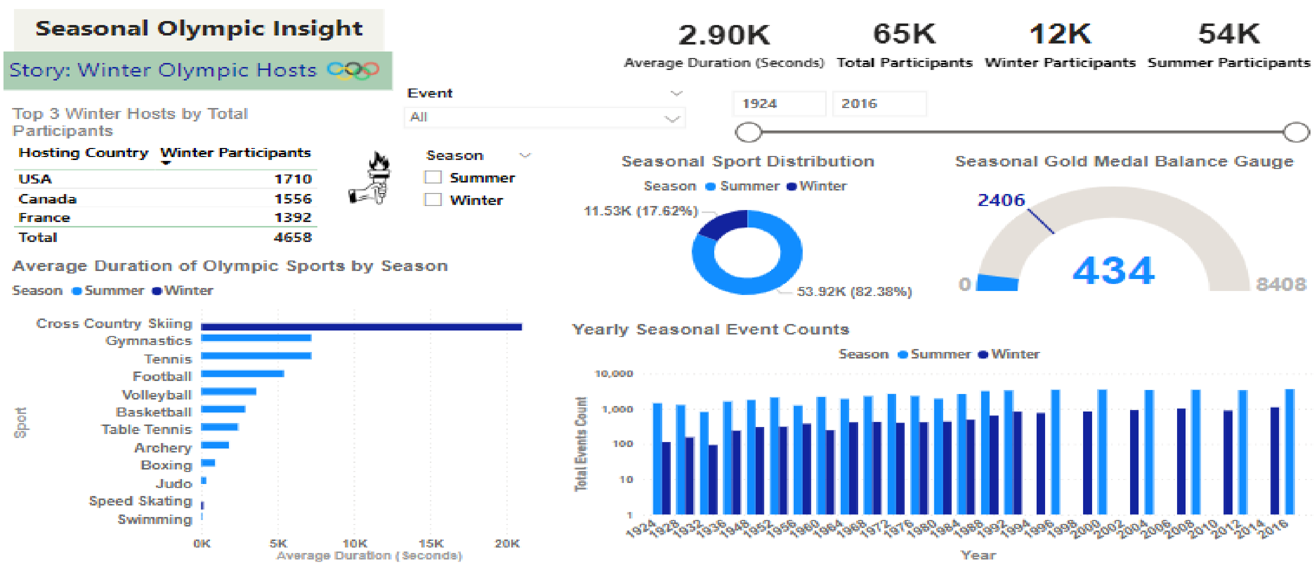

Figure 16: Seasonal Olympic Insight Dashboard

4.5 Results and Outcomes

4.5.1 Women's Participation in Olympic Dashboard

- Increase Women's Participation: The organization should actively work towards increasing the number of female participants in the Olympic Games, with the ultimate goal of achieving gender equality.
- Enhance Women's Events: Boost the number of women's events in the Olympic program to provide female athletes with more opportunities to compete at the highest level.
- Regional Focus: Targeted efforts should be made in regions with lower female participation rates to promote women's participation in the Olympics.

4.5.2 Jordan's Olympic Performance Analysis Dashboard

- Increase Participation: Jordan should focus on increasing the number of participants in the Olympic Games to maximize its chances of success.

- Encourage Youth Participation: Targeting individuals in their twenties and encouraging them to participate in the Olympics can help Jordan develop a new generation of athletes.
- Explore Winter Olympics: Jordan should explore opportunities for participation in the Winter Olympics to diversify its presence on the global Olympic stage.
- Support Athletes with Olympic Experience: Athletes with prior Olympic experience should be supported and encouraged to participate in subsequent games to leverage their expertise.
- Gender Diversity: Continue to encourage and promote the presence of Jordanian female athletes in Olympic events.

4.5.3 Seasonal Olympic Insight Dashboard

- Diversify Sports: Countries hosting the Winter Olympics may consider diversifying the range of sports to attract a broader audience and increase participant engagement.
- Promote Winter Sports: Focus on boosting the popularity of Winter Olympic sports to bridge the participation gap between Summer and Winter Games.
- Achieve Gold Medal Parity: Set the goal of achieving parity in gold medals between the Summer and Winter Olympics as a measure of overall success.

4.6 Conclusion

This project harnessed the potential of Microsoft Power BI to create three impactful Olympic data dashboards. These dashboards addressed Women's Olympic Participation, Jordan's Performance, and Seasonal Olympic Insights, providing actionable insights. Key takeaways from the project include the need to promote gender equality and increase female participation, enhance women's events, support Jordan's Olympic journey by expanding participation and exploring new opportunities like the Winter Olympics, and advising Winter Olympic hosts to diversify sports, promote winter events, and aim for gold medal parity. In essence, this project underscores the power of data visualization to inform decision-making and improve the Olympic experience.

5 Conclusion

In conclusion, these three projects collectively underscore the pivotal role of data analytics in addressing real-world challenges and opportunities. From database design that enhances user experiences, to analytics techniques that uncover hidden insights, and data visualization that informs strategic decisions, the significance of data analytics is undeniable. As data continues to proliferate, the ability to harness its potential remains a critical skill, enabling individuals and organizations to thrive in our data-driven era. This report serves as a testament to the transformative power of data analytics across diverse domains, emphasizing its relevance and impact in the contemporary technological landscape.

References

1) scikit-learn: machine learning in Python — scikit-learn 1.3.1 documentation
2) NumPy Documentation
3) 3.12.0 Documentation (python.org)
4) Microsoft SQL documentation - SQL Server | Microsoft Learn