Data Journey:

SQL, Anomaly Categories Detection, and Power BI

Mays Moh’d Al-Fasfous

3

2

1

chapter

Proiect

Presentation

0 1

Online Ticket Booking Database System

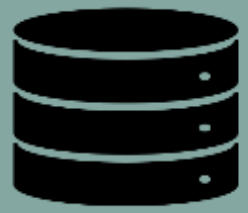
3

2

chapter

Projects

By using SQL Server



1

chapter

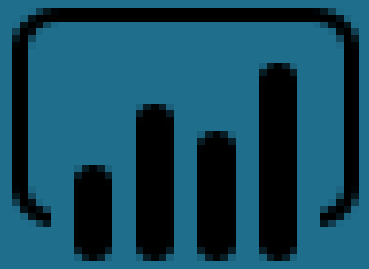
Projects

Descriptive and Predictive Analytics for Anomaly Categories

By using Python

Exploring Olympic Data: Visualizations and Insights

By using Power BI



chapter Projects

1

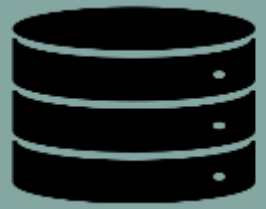
2

3

0 1

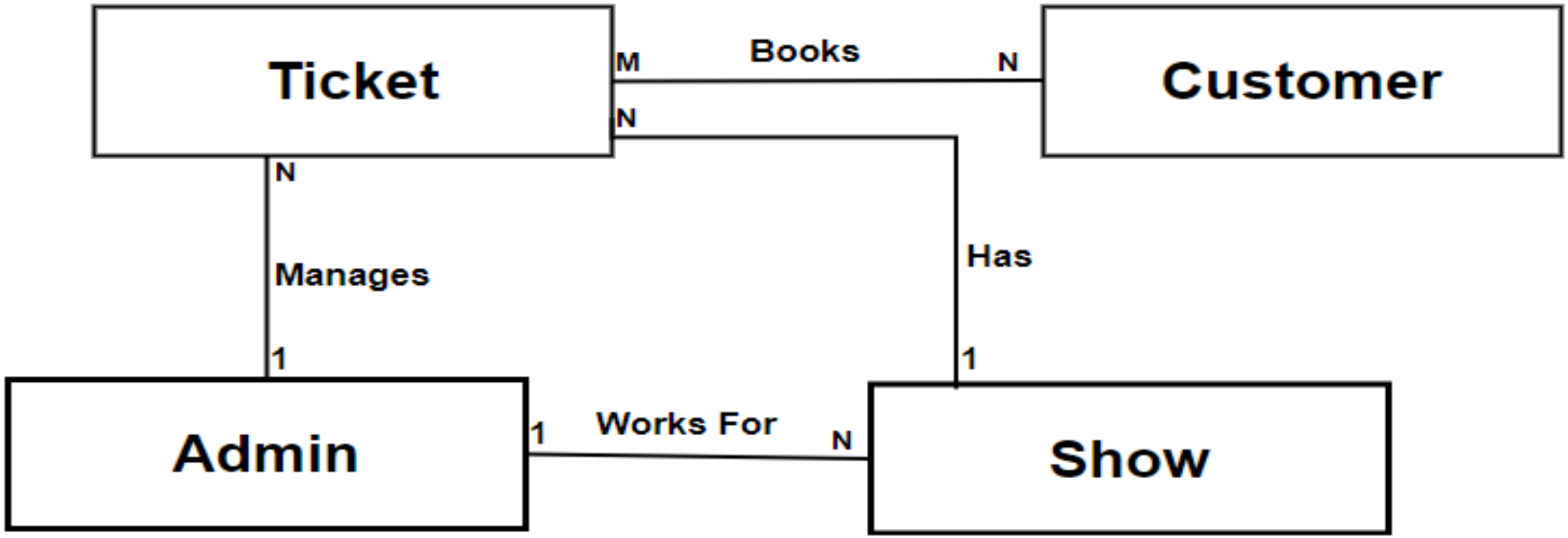
Online Ticket Booking Database System

By using SQL Server

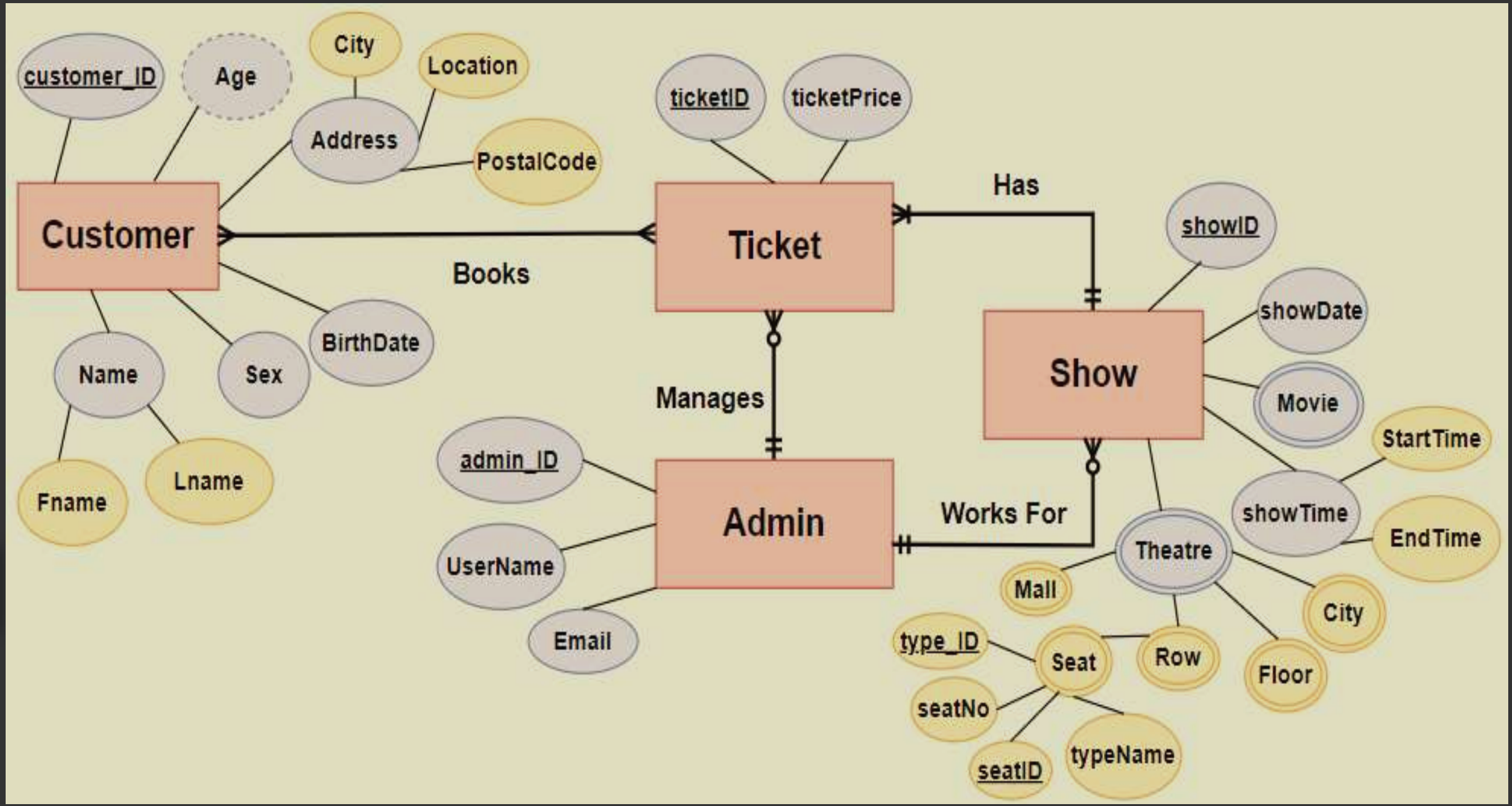


Conceptual Model for the

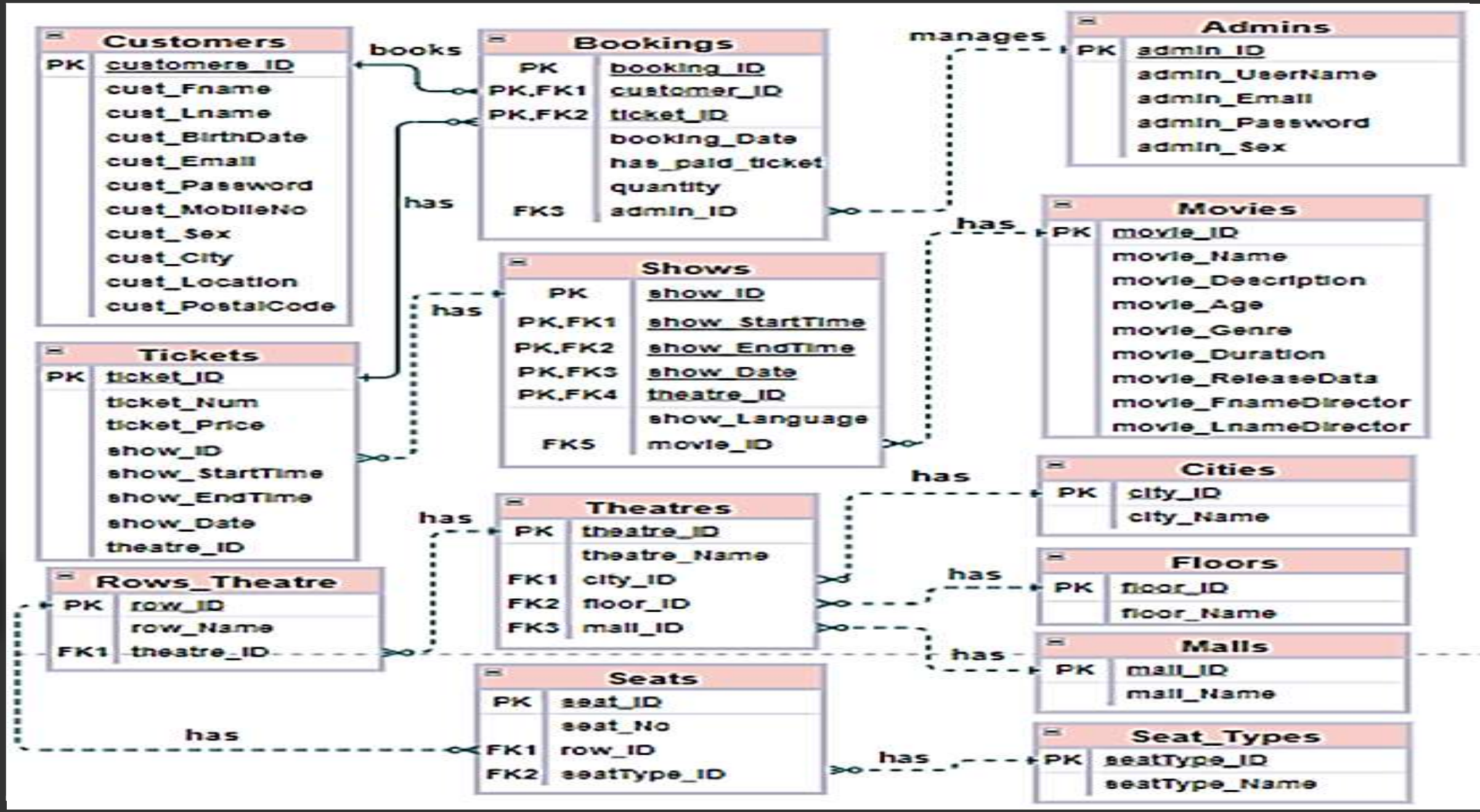
Online_Movie_Booking System DataBase



Conceptual Model with attributes

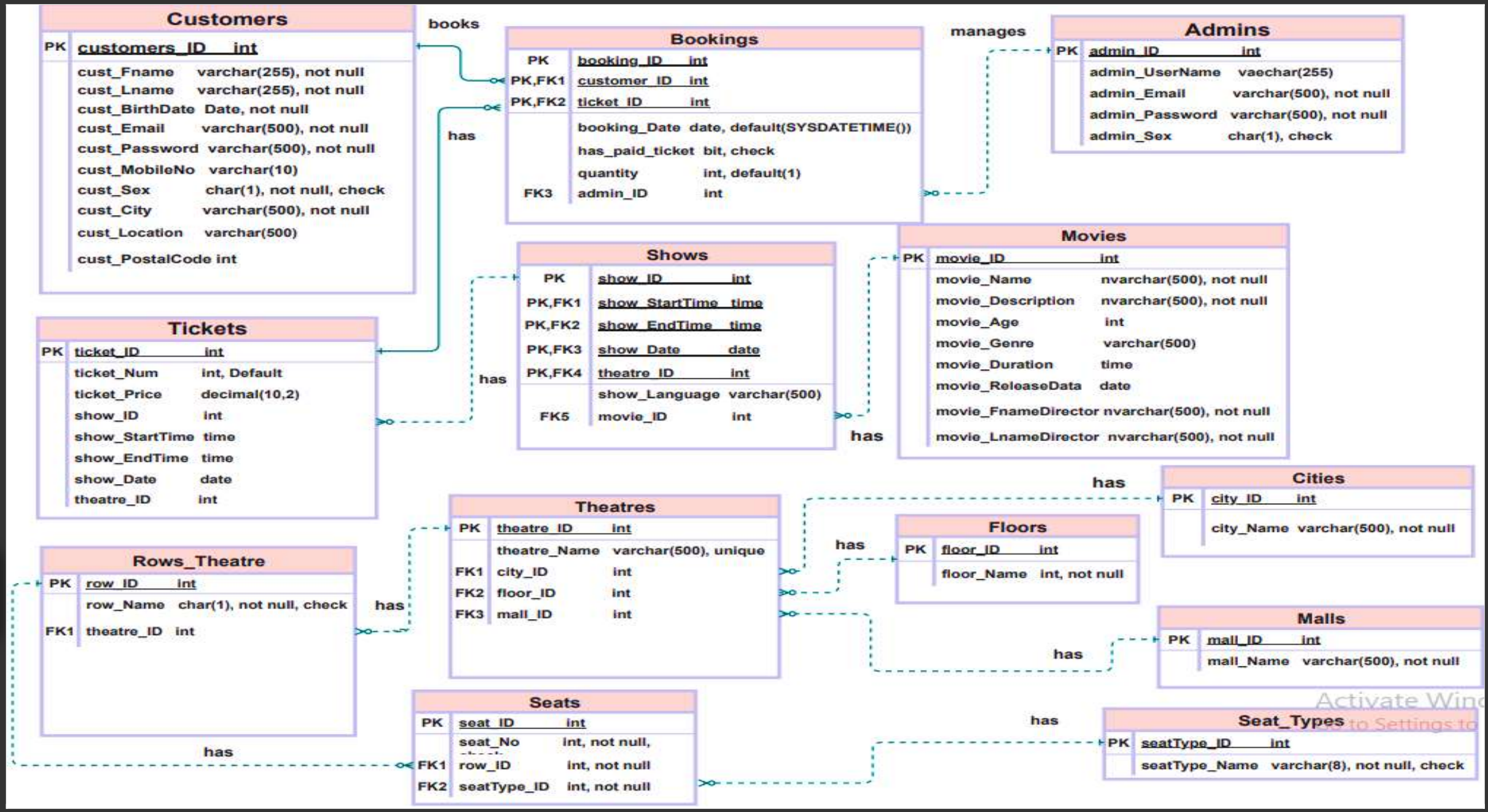


Logical Model for Online Movie Booking System Database



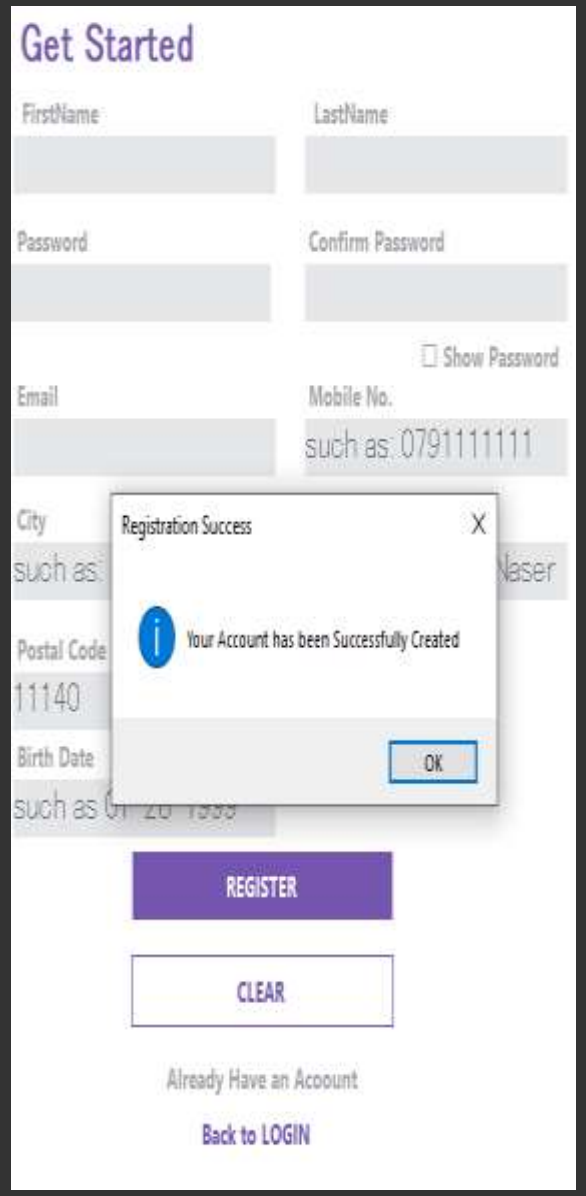
1

Physical Model for Online Movie Booking System Database



1

Integrating the Database with a Simple User Interface




```
-- DQL
select * from Customers;
```

120 %

Results | Messages

	customer_ID	cust_Fname	cust_Lname	cust_BirthDate	cust_Email	cust_Password	cust_MobileNo	cust_Sex	cust_City	cust_Location	cust_PostalCode
1	1	Saleh	Abdullah	1990-05-15	saleh.abdullah@gmail.com	password1	1234567890	M	Amman	Rainbow Street	11110
2	2	Layla	Mohammed	1985-08-20	layla.mohammed@gmail.com	password2	9876543210	F	Irbid	Abu Bakr Street	21110
3	3	Ahmed	Ali	1993-03-10	ahmed.ali@gmail.com	password3	5555555555	M	Zarqa	Al-Zaytoon Street	13100
4	4	Fatima	Hassan	1988-12-05	fatima.hassan@gmail.com	password4	1111111111	F	Aqaba	Al-Hussein Street	77110
5	5	Omar	Khalid	1991-09-25	omar.khalid@gmail.com	password5	2222222222	M	Karak	Al-Mutasim Street	66110
6	6	Mona	Hamed	1987-07-15	mona.hamed@gmail.com	password99	9999999999	F	Tafilah	Al-Ahmad Street	66111
7	7	Hassan	Ibrahim	1992-02-03	hassan.ibrahim@gmail.com	password100	8888888888	M	Maan	Al-Hussein Street	72110
8	8	yasmeen	khaled	1998-01-02	yasmeenkhaled@gmail.com	12345	0788888888	F	Amman	Jabel AlNaser	11140

1

Descriptive and Predictive Analytics for Anomaly Categories

By using Python

Descriptive Analytic Techniques

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

- Calculate thresholds by IQR:

upper_outlier_threshold = Q3 + (1.5 *IQR_duration)

lower_outlier_threshold = Q1 - (1.5 * IQR_duration)

Flow_Duration:

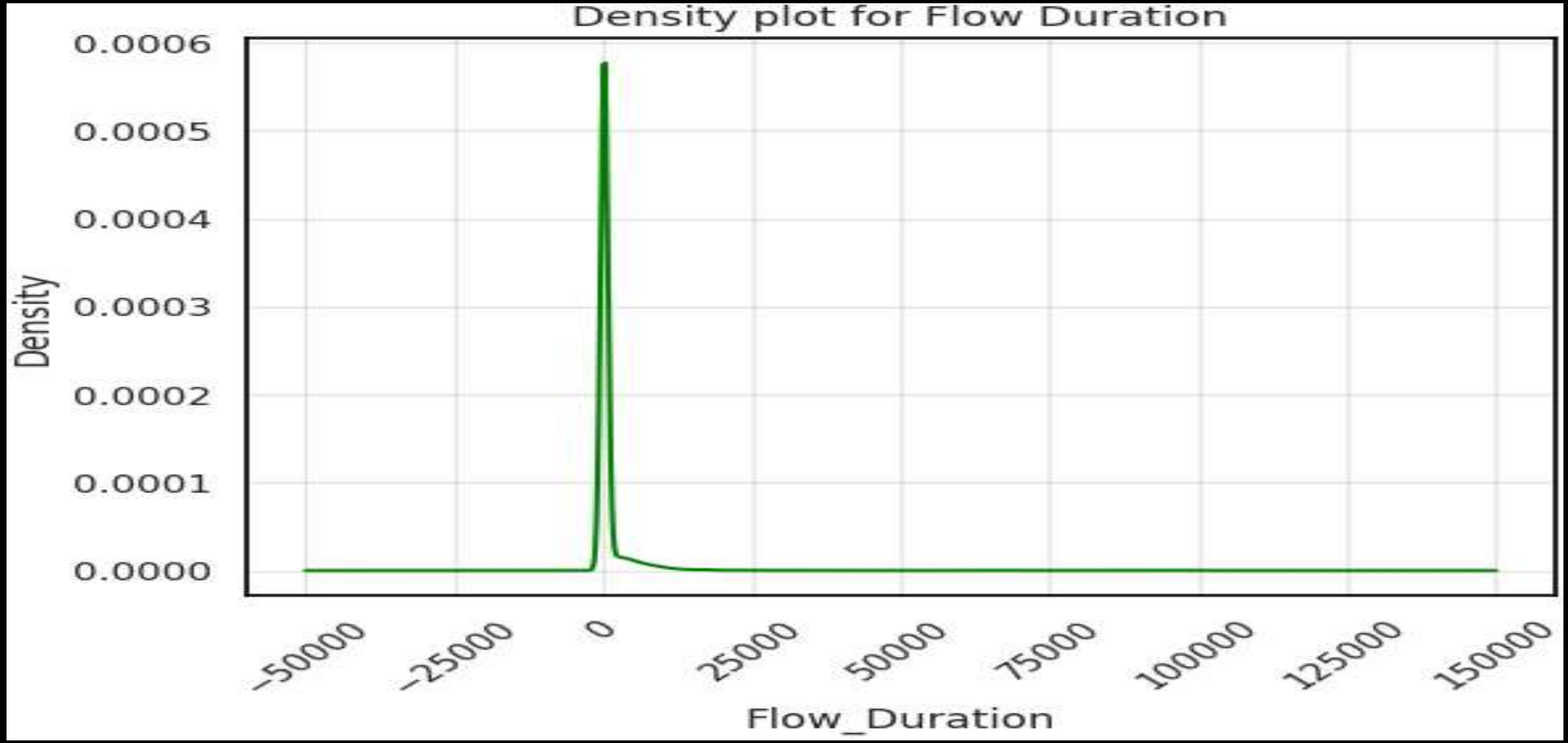
- Upper outlier = 481.0
- Lower outlier = - 296.0

Idle_Mean:

- Upper outlier = 326.5
- Lower outlier = - 141.5

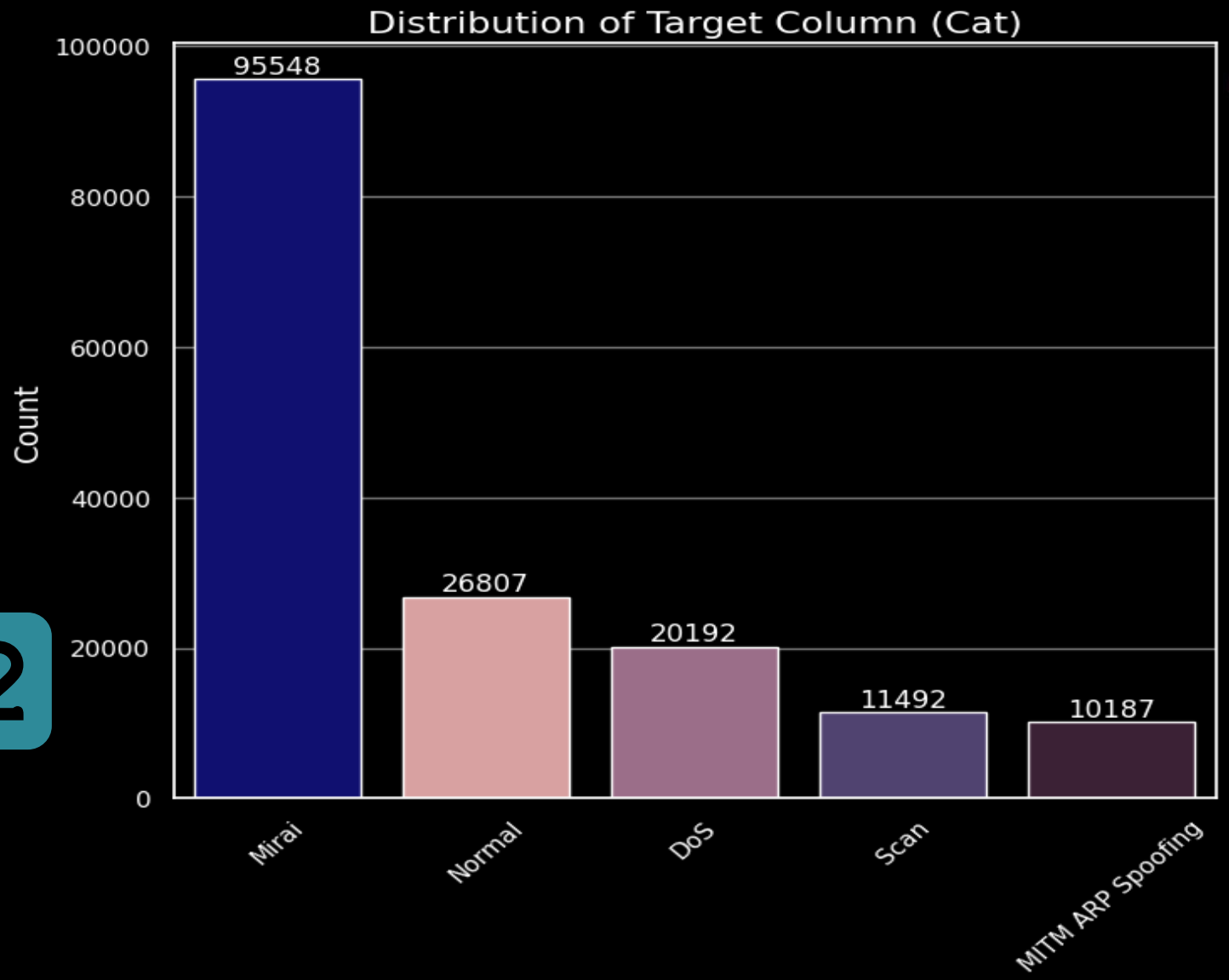
Descriptive Analytic Techniques

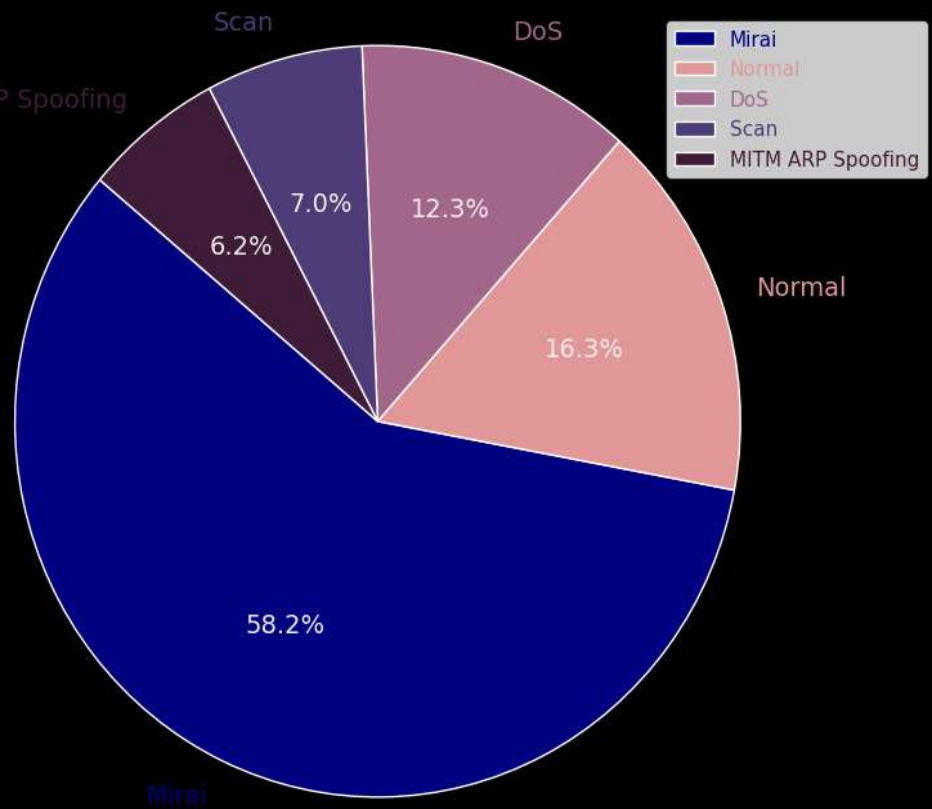
Outlier detections / Density plot



Descriptive Analytic Techniques

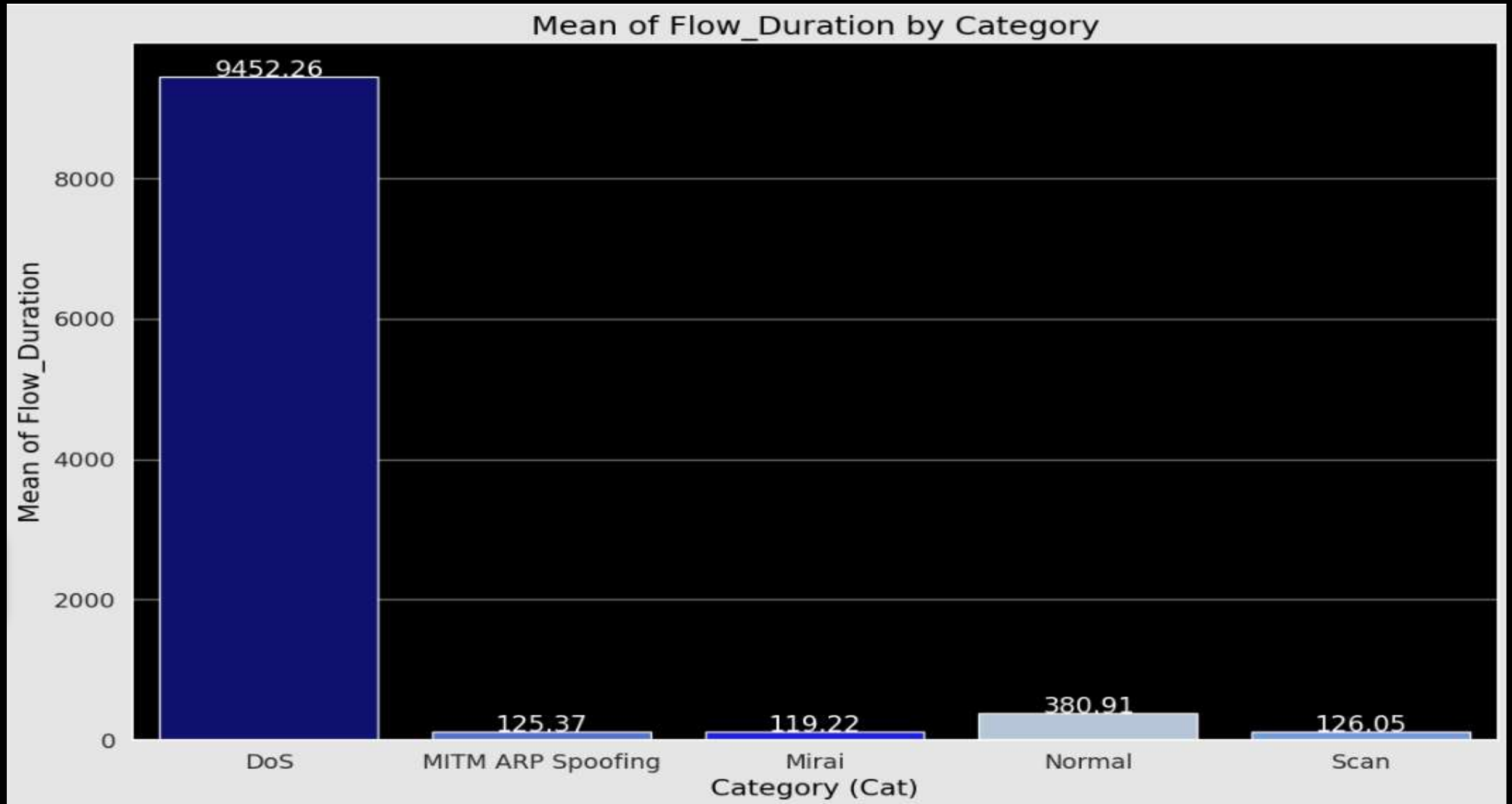
Descriptive Analytic Techniques for categorical columns (‘Cat’ column)



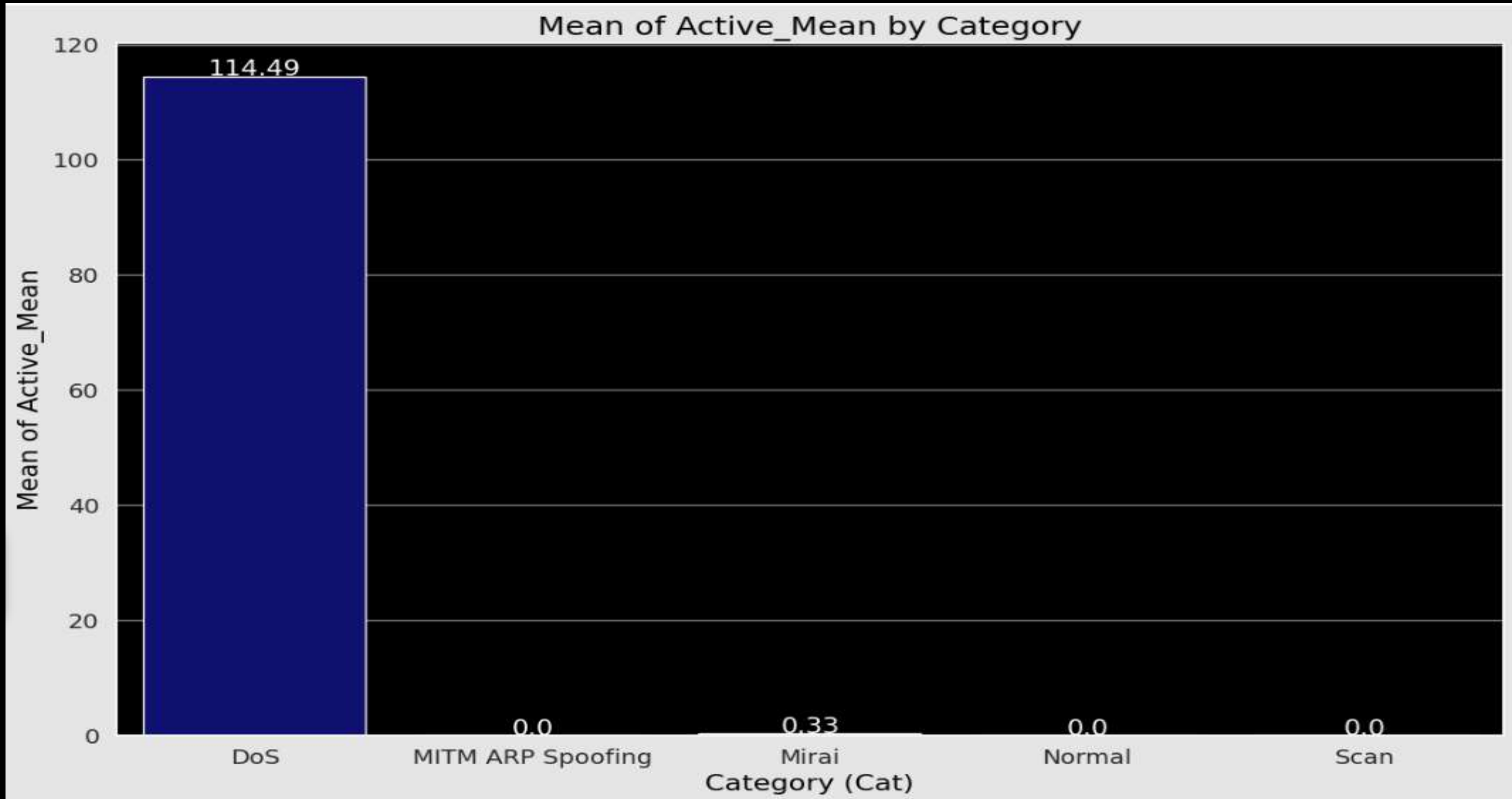


2

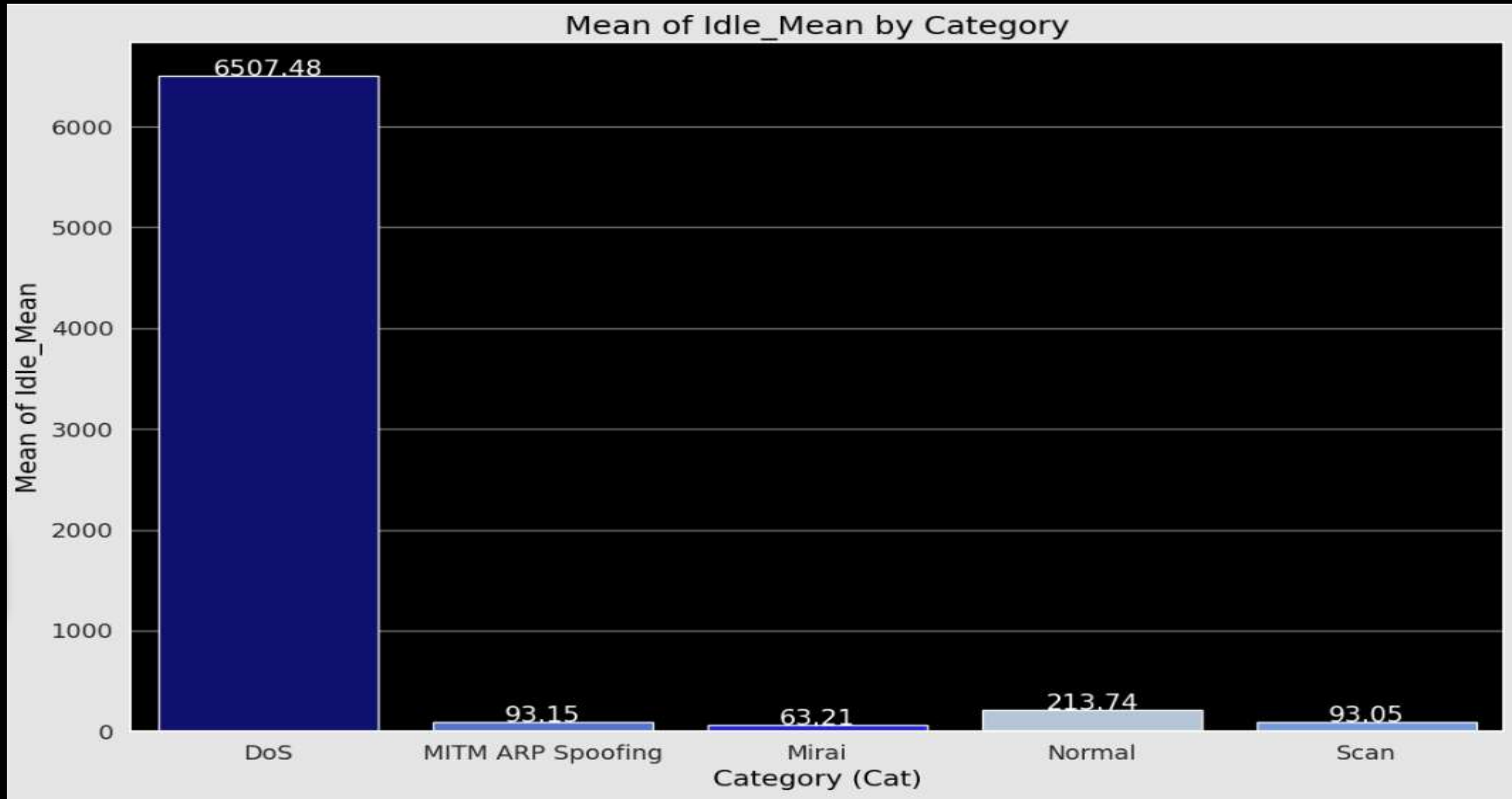
Descriptive Analytic - Insight



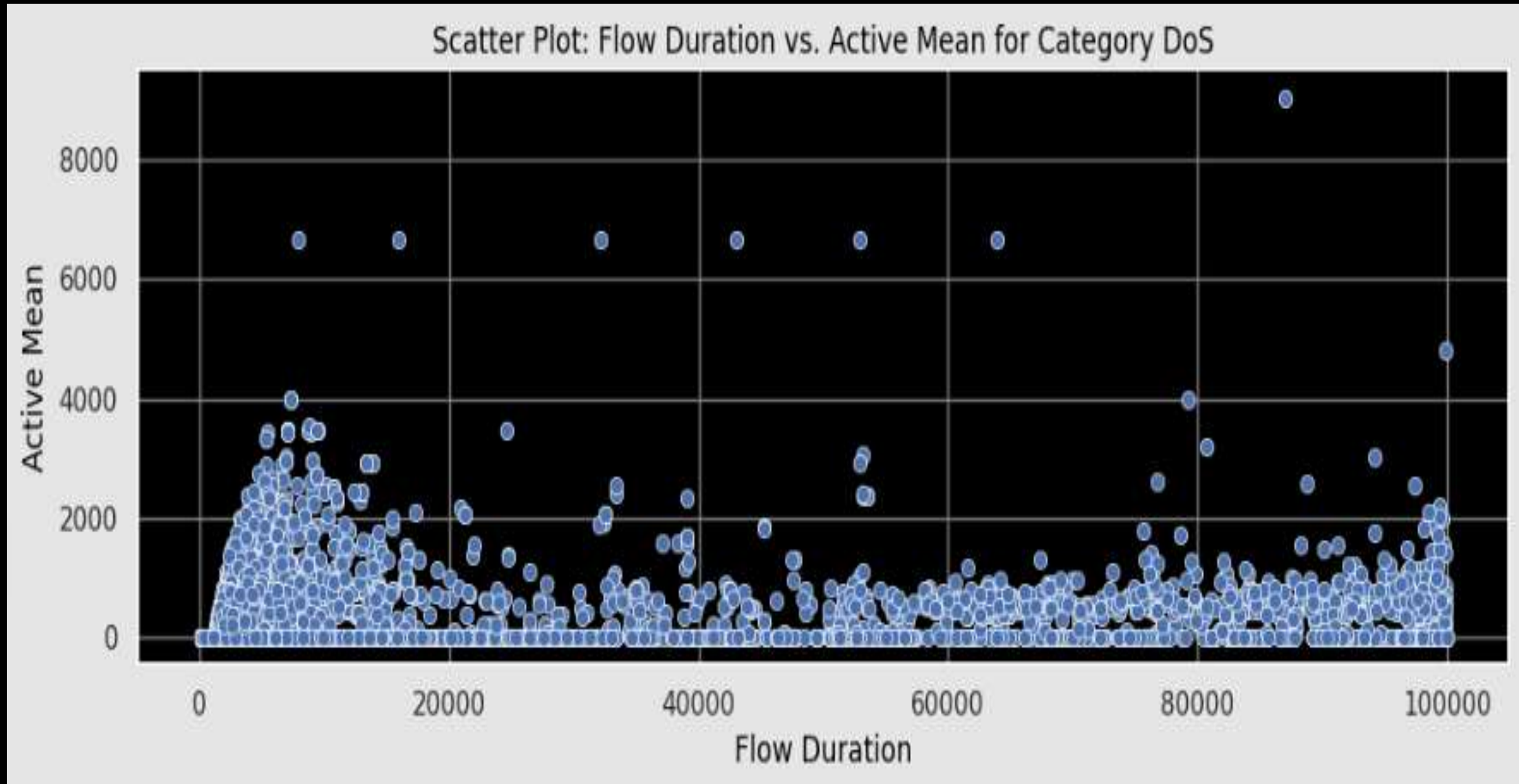
Descriptive Analytic - Insight



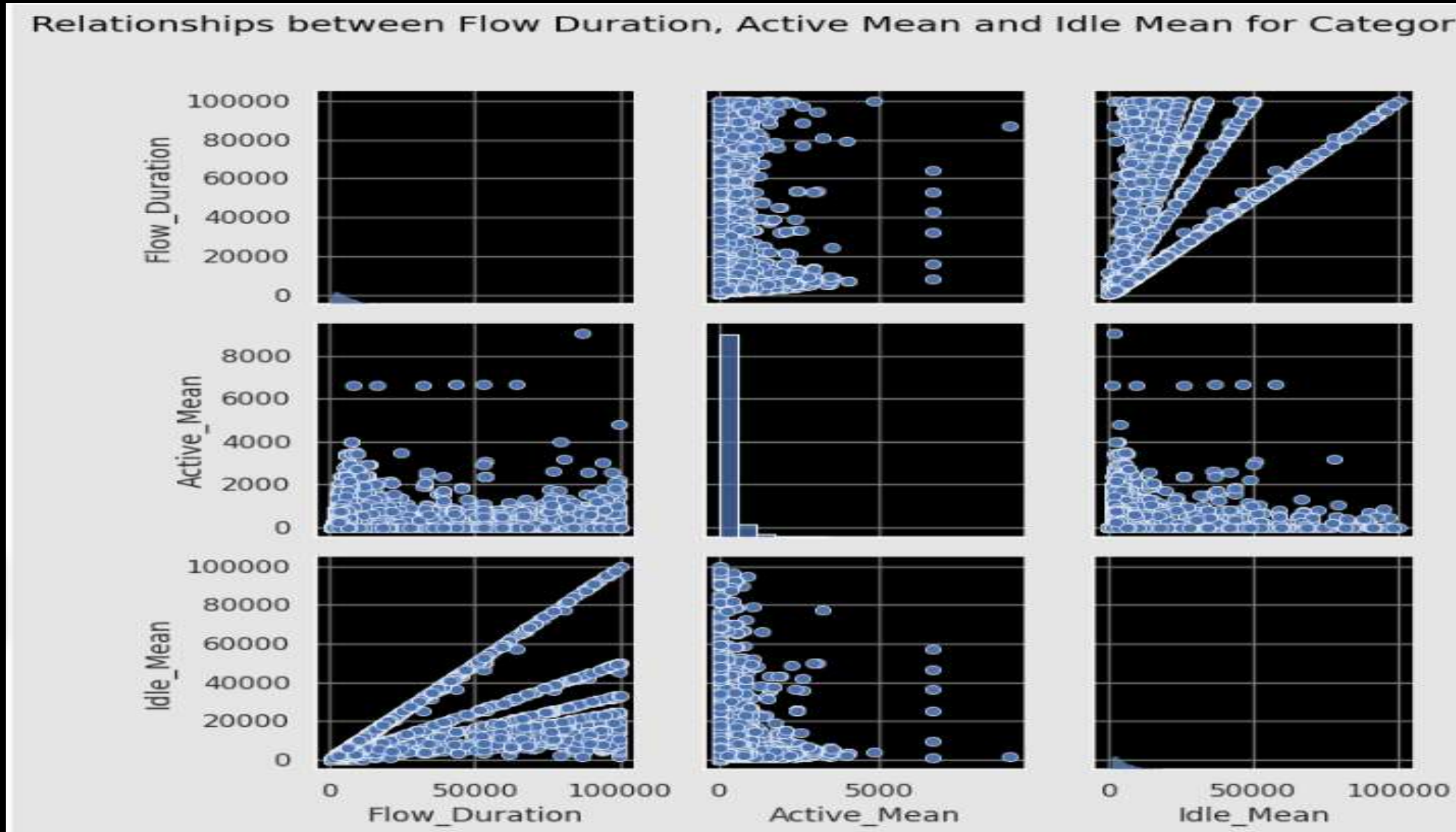
Descriptive Analytic - Insight



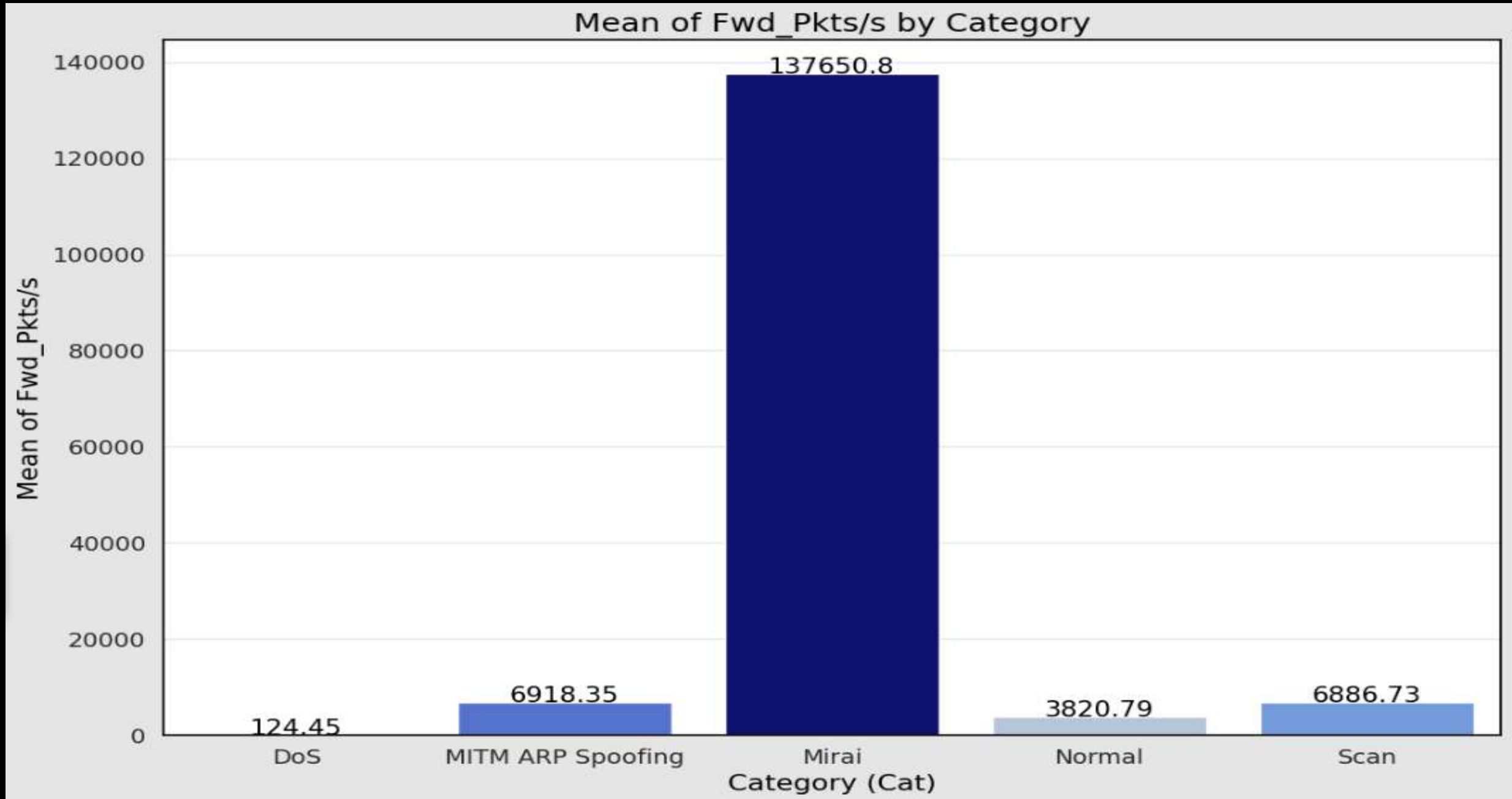
Descriptive Analytic - Insight



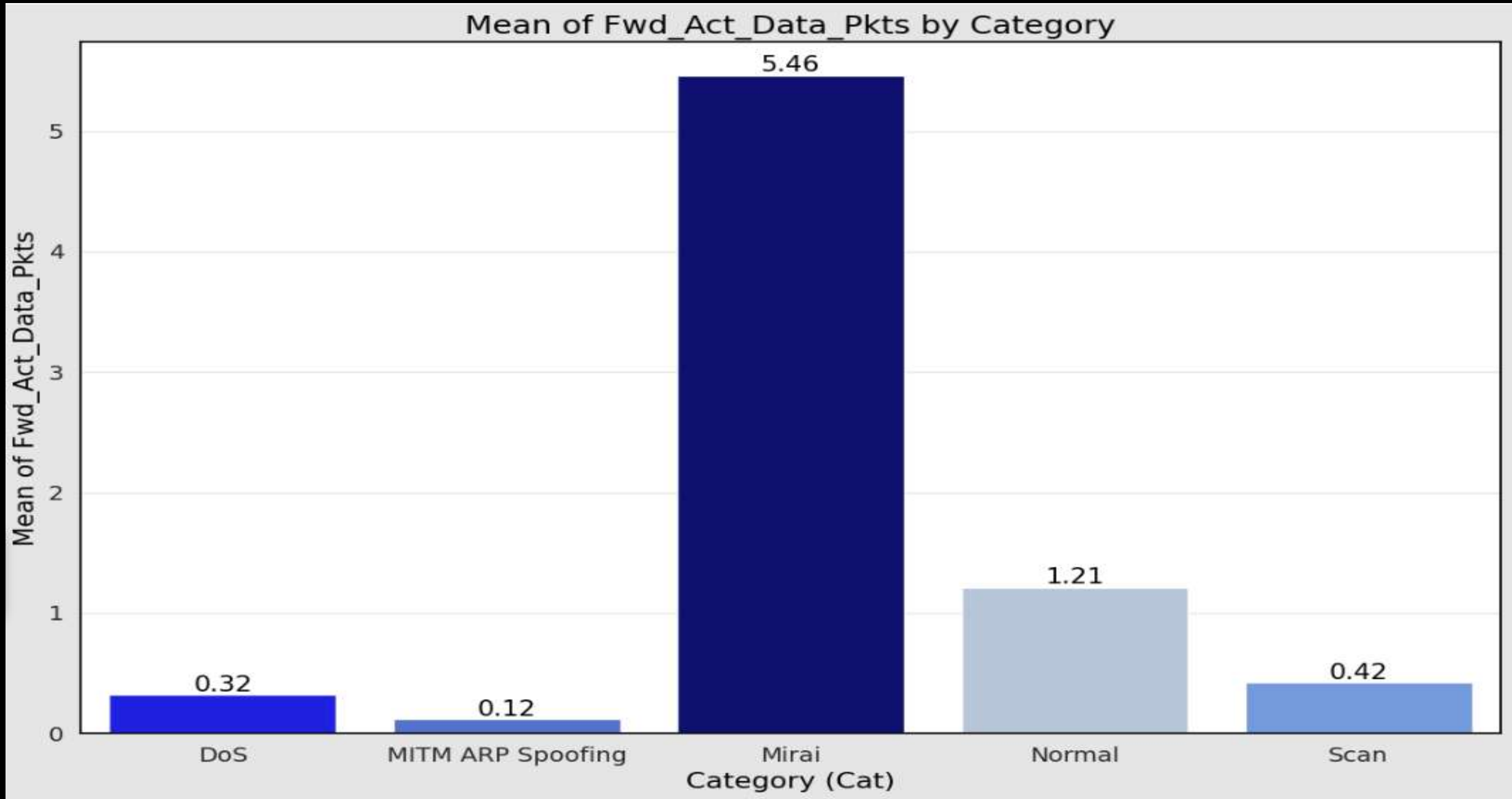
Descriptive Analytic - Insight



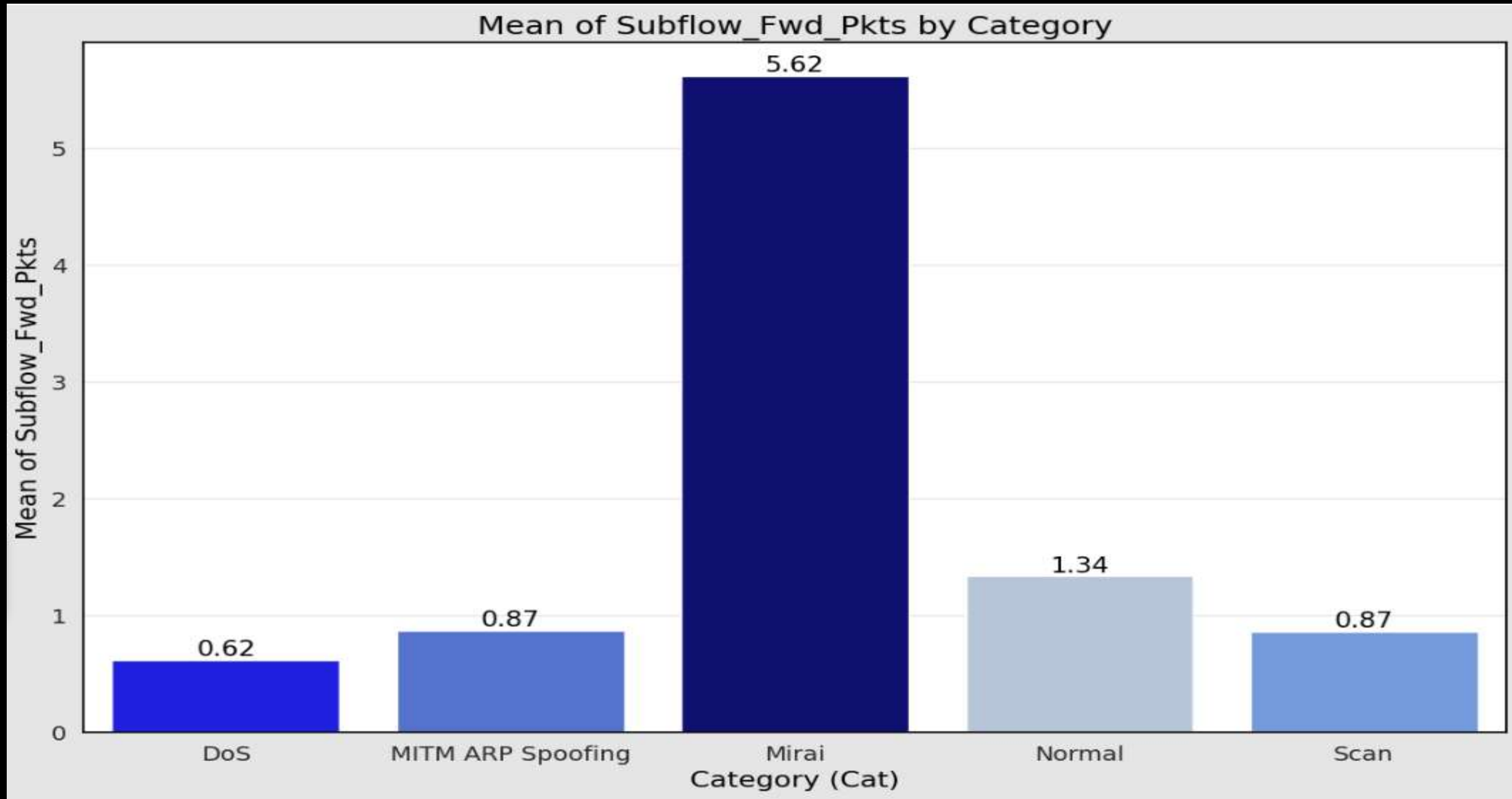
Descriptive Analytic - Insight



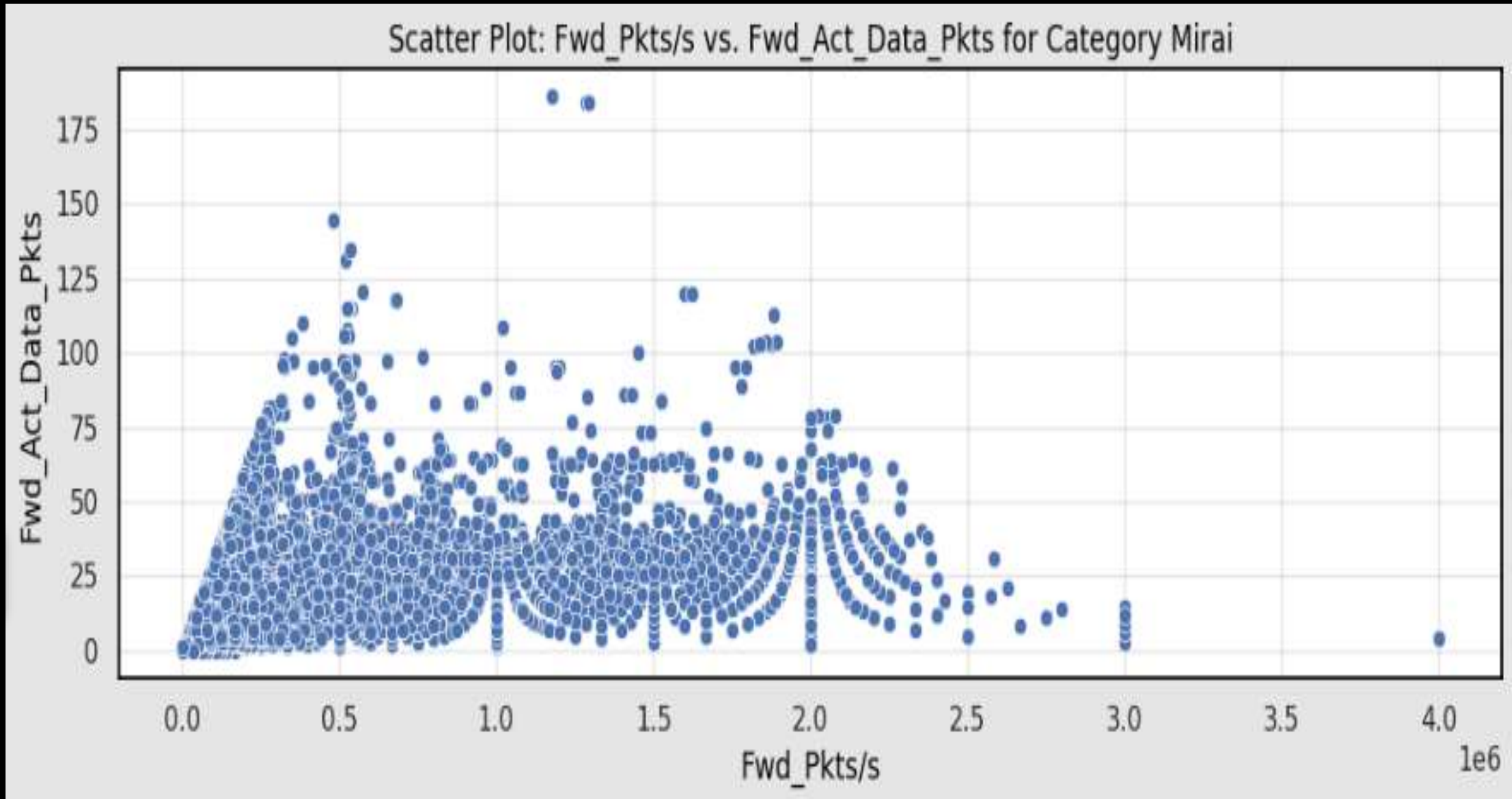
Descriptive Analytic - Insight



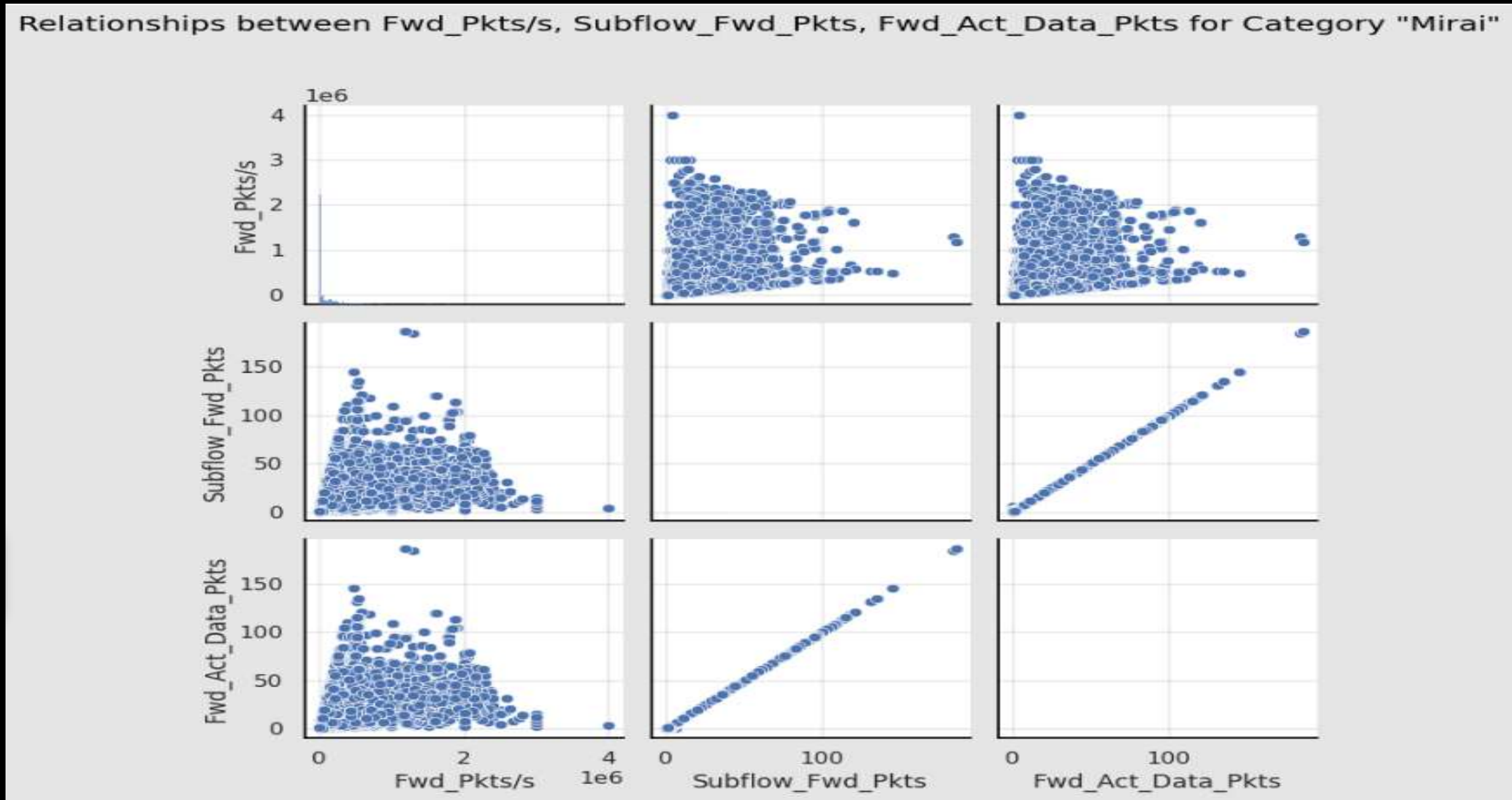
Descriptive Analytic - Insight



Descriptive Analytic - Insight



Descriptive Analytic - Insight



Key Insights Affecting Decisions

- ✓ In the context of the 'DoS' category, three key attributes, 'Flow_Duration,' 'Active_Mean,' and 'Idle_Mean,' have shown significant relationships with the classification of incidents into the 'DoS' category. This suggests their importance in the feature selection and model-building processes to enhance the accuracy of 'DoS' category classification.

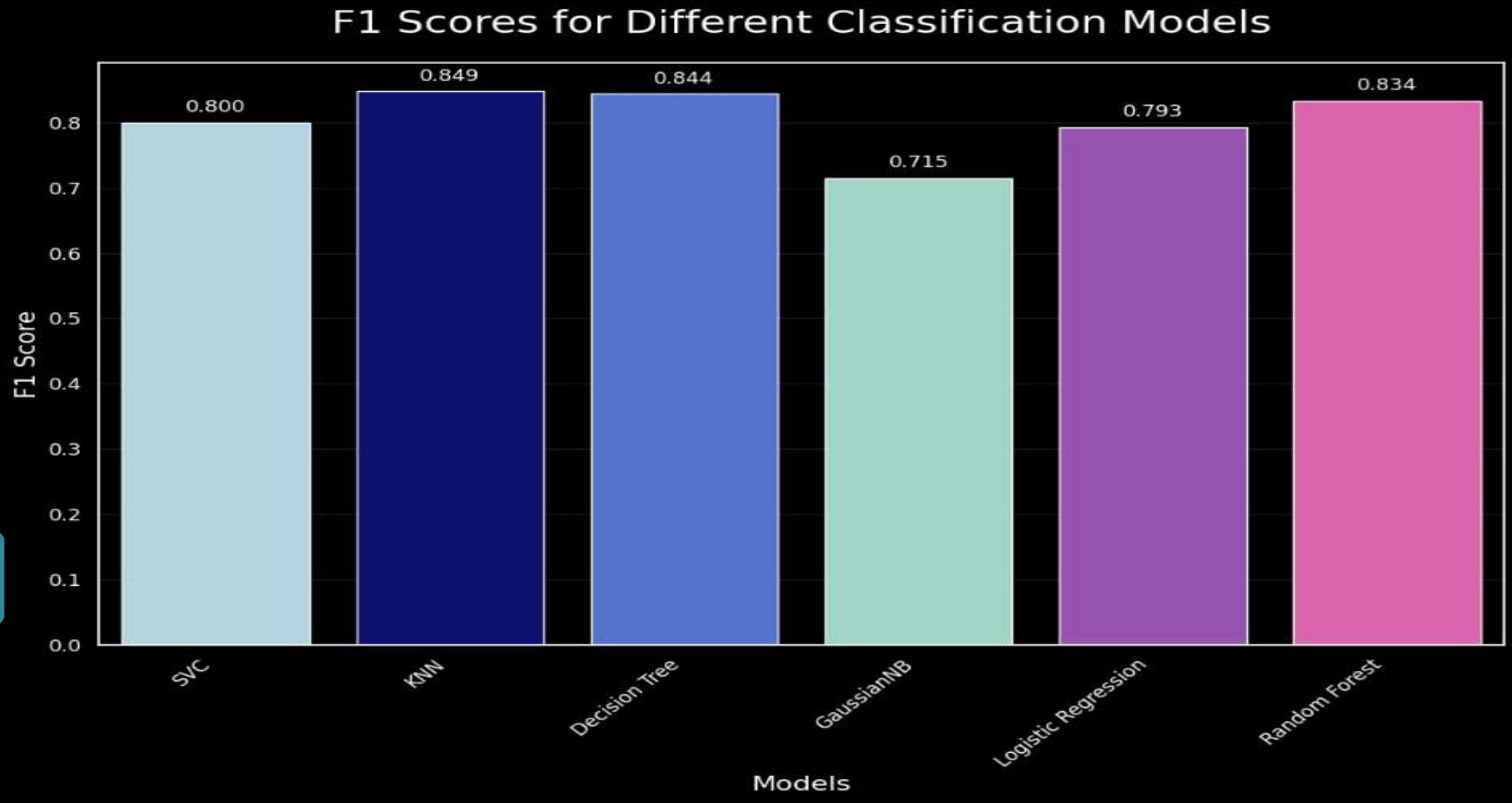
- ✓ In the context of the 'Mirai' category, three key attributes, 'Fwd_Pkts/s,' 'Subflow_Fwd_Pkts,' and 'Fwd_Act_Data_Pkts,' have shown significant relationships with the classification of incidents into the 'Mirai' category. This suggests their importance in the feature selection and model-building processes to enhance the accuracy of 'Mirai' category classification.

Predictive Analytic

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

Predictive Analytic



Predictive Analytic

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

Predictive Analytic

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

Predictive Analytic

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

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

Predictive Analytic

Comparison of KNN and DT Performance

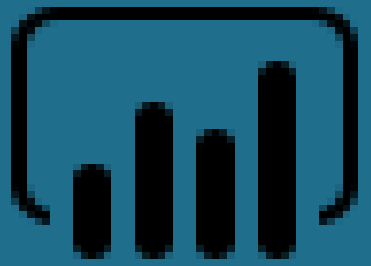
	Models	Accuracy	Precision	Recall	F1-Score
0	DT With With SelectFwe	0.902417	0.894703	0.902417	0.896636
1	KNN With VarianceThreshold	0.885403	0.875026	0.885403	0.877920

2

- The best performing predictive analytics model is the Decision Tree (DT) with 'SelectFwe' feature selection.

Exploring Olympic Data: Visualizations and Insights

By using Power BI



Women's Participation in Olympic Dashboard

Story: Equality in Olympic Participation

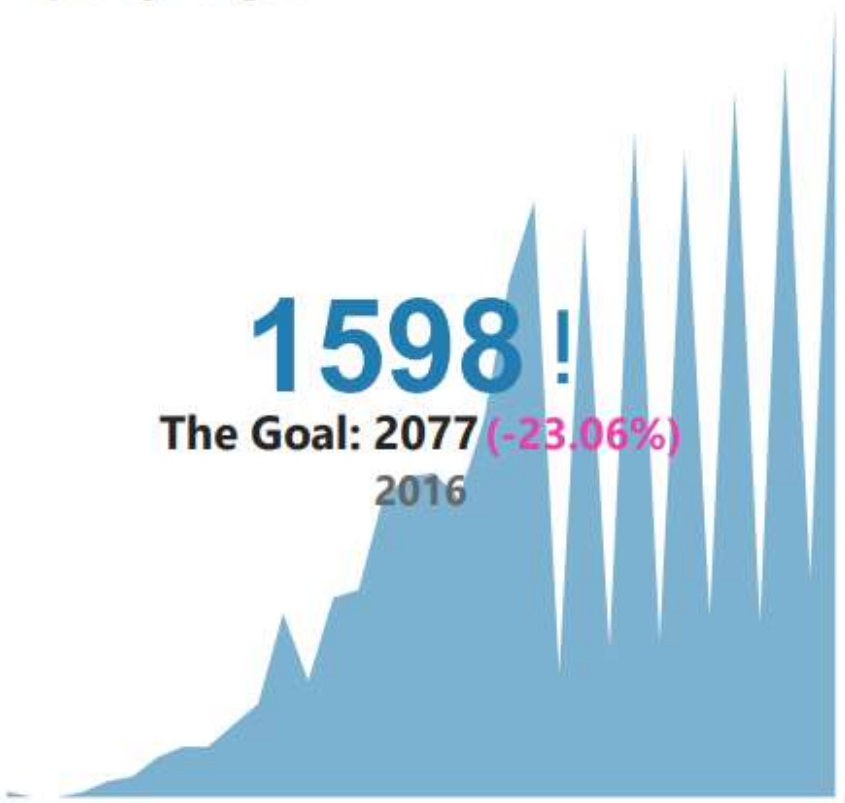
Total Participants	Male Participants	Female Participants
70K	52K	18K

Region	Female_Participants
USA	1451
France	1041
Germany	1001
Total	3493

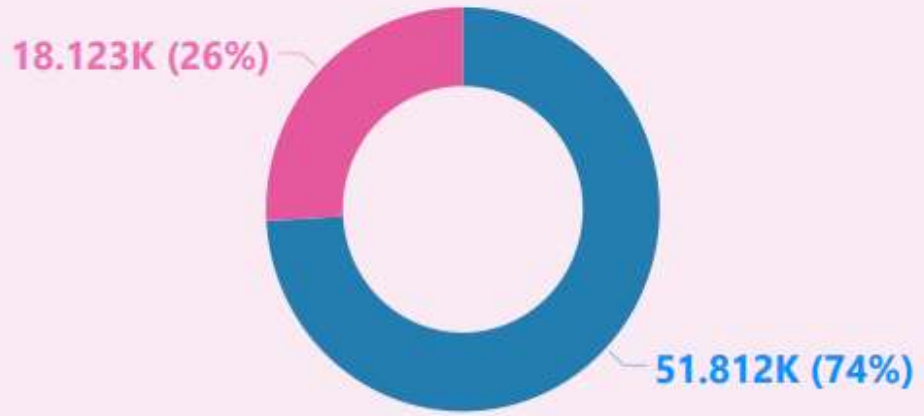
Region

All

Equality Target



Gender Distribution



1900 2016

Equality in Participation | Events and Regional

Activate Windows
Go to Settings to activate Windows.

Women's Participation in Olympic Dashboard

Story: Women's Olympic Events Analysis and Regional Insight

210
WomensEventCount

439
MensEventCount

87
MixedEventCount

69935
Count of Event

Team	Total Medals	Gold	Silver	Bronze
United States	495	246	141	108
Australia	145	48	49	48
Germany	132	43	47	42
Total	772	337	237	198

Top 3 Female's Teams

Olympic Events Distribution Over Time

Gender Distribution in Mixed Events

Regional Distribution of Female Participants

1900 2016

Event Name

- Alpine Skiing Men's Combined
- Alpine Skiing Men's Downhill
- Alpine Skiing Men's Giant Slalom
- Alpine Skiing Men's Slalom
- Alpine Skiing Men's Super G
- Alpine Skiing Women's Combined
- Alpine Skiing Women's Downhill
- Alpine Skiing Women's Giant Sl...

Equality in Participation

Events and Regional

Activate Windows
Go to Settings to activate Windows.

Key Insights Affecting Decisions

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 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.

Jordan's Olympic Performance Analysis

Story: Analyzing Jordan's Olympic Journey

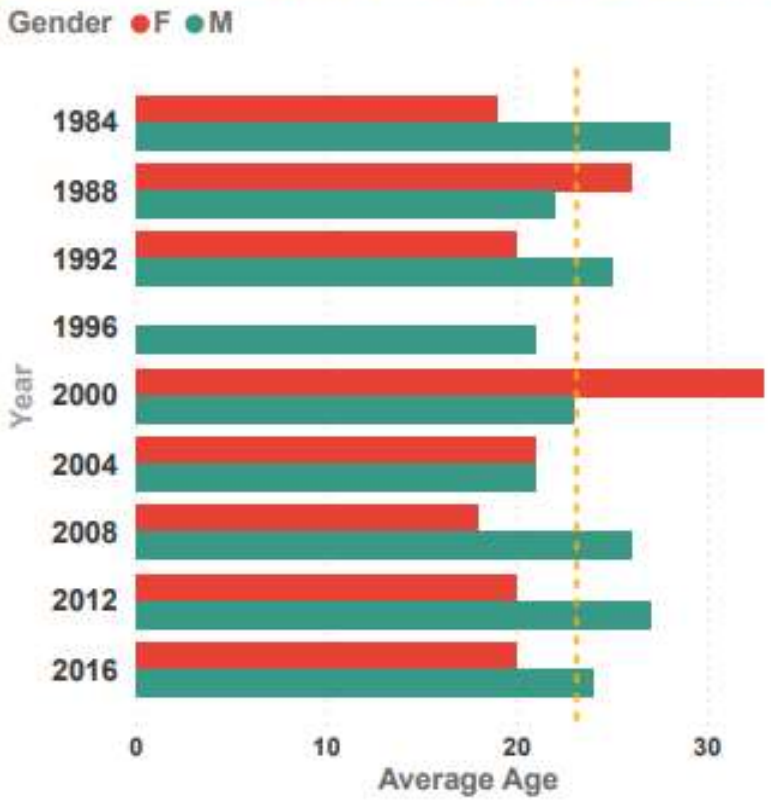
47
Total Participants

24
Average Age

1
Total Medals

JORDAN

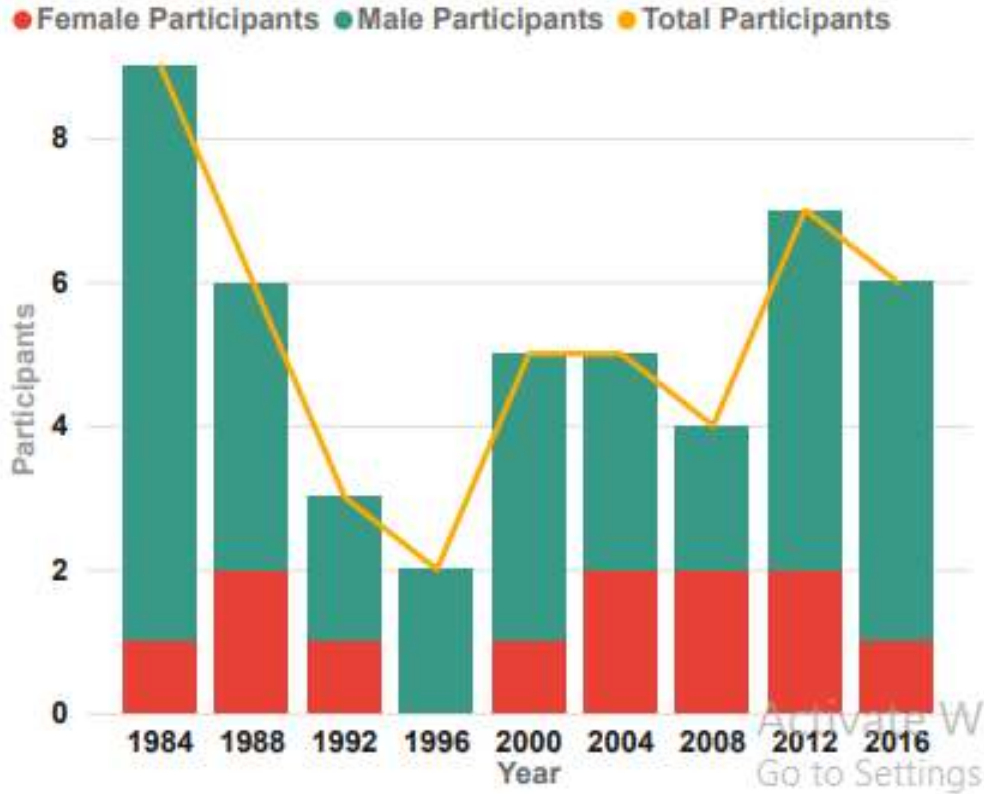
Average Age of Jordanian Olympic Participants



Medal
All

Season
All

Jordan's Olympic Participation Trends Over the Years



Name	Sport	Games	Medal	Total Participants
Ahmad Abughaush	Taekwondo	2016 Summer	Gold	1
Khader Ghetrich Baqlah	Swimming	2016 Summer	No Medal	1
Lawrence Fanous	Triathlon	2016 Summer	No Medal	1
Total				47

Activate Wind
Go to Settings to

Key Insights Affecting Decisions

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 and explore Winter Olympics.
- ✓ 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.
- ✓ Support Athletes with Olympic Experience: Should be supported and encouraged to participate in subsequent games to leverage their expertise.
- ✓ Continue to encourage and promote the presence of Jordanian female athletes in Olympic events.

Seasonal Olympic Insight

Story: Winter Olympic Hosts

2.90K	65K	12K	54K
Average Duration (Seconds)	Total Participants	Winter Participants	Summer Participants

Event: All

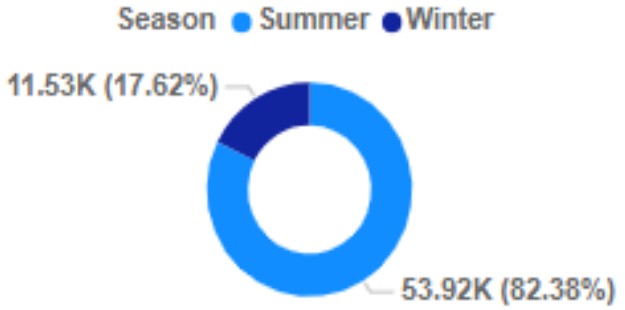
1924 — 2016

Top 3 Winter Hosts by Total Participants

Hosting Country	Winter Participants
USA	1710
Canada	1556
France	1392
Total	4658

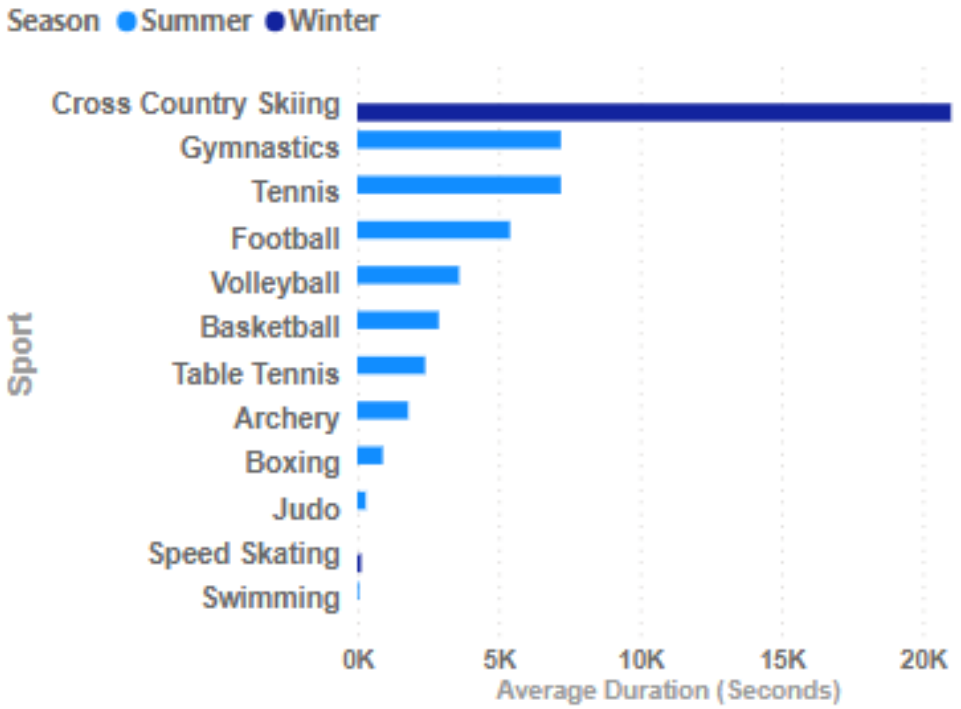
Season
- [ ] Summer
- [ ] Winter

Seasonal Sport Distribution

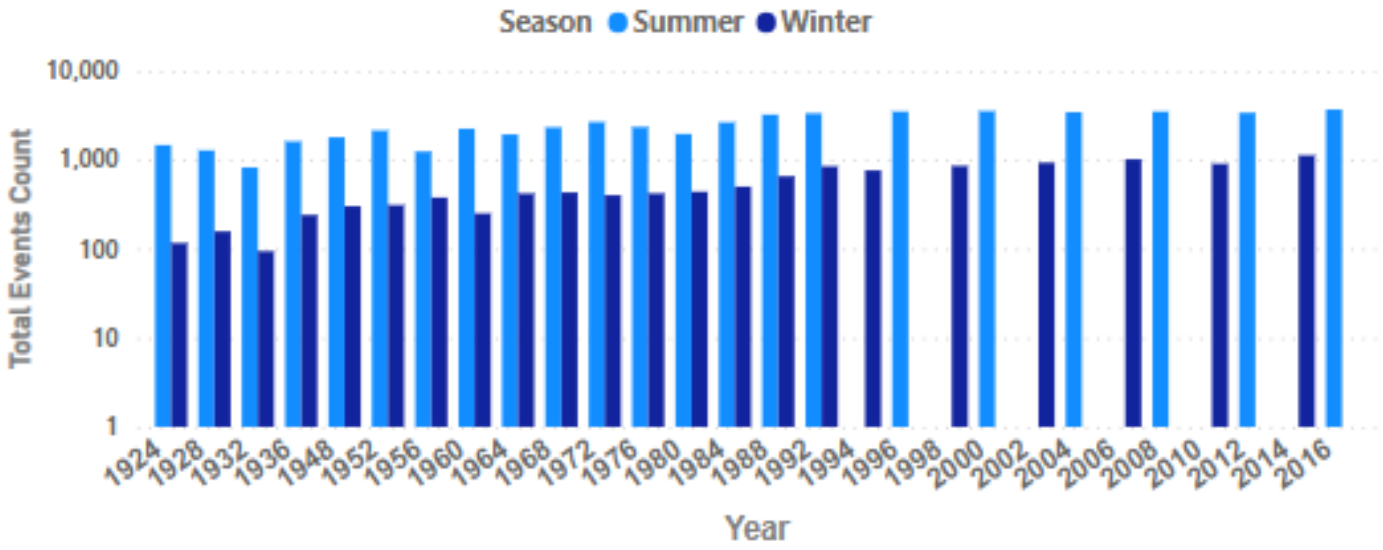


Seasonal Gold Medal Balance Gauge

Average Duration of Olympic Sports by Season



Yearly Seasonal Event Counts



3

Key Insights Affecting Decisions

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.

Thank you

Mays Moh’d Salih Al-Fasfous

maysmalfasfous@gmail.com

http://www.linkedin.com/in/maysalfasfous