SPRINT 1 DOCUMENT

PROJECT TITLE: SPACE TRAFFIC DENSITY PREDICTION

Team Members:

1. Himaja S
2. Tarunaa A C
3. Karanbir Singh
4. Shravani R S
5. Anushka Dutta

1. SPRINT OBJECTIVE

To perform exploratory data analysis (EDA), pre-processing, and feature engineering to gain insights into the dataset and set the groundwork for predictive modeling of space traffic density.

2. SPRINT OVERVIEW

Duration: 7-Nov-24 to 18-Nov-24

Focus Areas:

- Understanding the problem statement.
- Gaining insights into the dataset.
- Conducting exploratory data analysis (EDA).

3. DELIVERABLES

3.1 Problem Statement

- Addressed the need for a reliable space traffic density prediction system for collision avoidance and mission planning.
- Identified key factors influencing traffic density: orbital location, object type, and time of day.
- Developed a foundational understanding of project expectations and objectives.

Expected Outcomes Delivered:

- Clear understanding of the problem.
- Established scope and requirements for predictive modeling.

3.2 Dataset Exploration

- Attributes analyzed:
    - **Timestamp:** Traffic recording time.
    - **Location:** Orbital zones like Lagrange Points, Low Earth Orbit, Mars Transfer Orbit.
    - **Object_Type:**Categories like Space Stations, Satellites, Probes, and Space Debris, Manned Spacecraft and Asteroid Mining Ship.

    - **Traffic_Density:** Numerical representation of space traffic volume.
    - **Peak_Time:** Hourly peak traffic observations.
- Dataset summary and insights:
    - No missing values.
    - Nearly symmetric Traffic_Density distribution (skewness = 0.092).
    - Data types: Numerical and categorical attributes.

3.3 Exploratory Data Analysis (EDA)

Data Insights:

1. **Univariate Analysis:**
    - Distribution and characteristics of individual features.
    - Visualizations: Histograms, Box Plots, and Violin Plots for Traffic_Density.
2. **Bivariate Analysis:**
    - Examined Traffic_Density trends across locations and object types using box and violin plots.
    - No significant association between object type and location (Chi-square p-value = 0.3583).
3. **Time Series Analysis:**
    - Observed traffic density trends over time.
4. **Additional Analysis:**
    - Identified high-risk zones and bottlenecks for specific object types.
    - Highlighted peak traffic patterns by object type and location.

4. KEY METRICS

Metric	Value
Total Records in Dataset	1000
Skewness of Traffic_Density	0.092 (Nearly symmetric)
ANOVA Test (p-values for Object_Type and Location)	0.464, 0.339 (Not significant)
Chi-Square Test (Object_Type vs Location)	p-value = 0.3583

5. KEY LEARNINGS

- Gained a deeper understanding of orbital dynamics and traffic patterns.
- Learned advanced visualization techniques to uncover patterns in data.
- Importance of robust feature engineering for categorical variables.

6. NEXT STEPS

- Refine feature engineering techniques for predictive modeling.
- Design and implement machine learning models for space traffic prediction.
- Explore advanced time series forecasting methods for real-time prediction.

7. VISUALIZATIONS

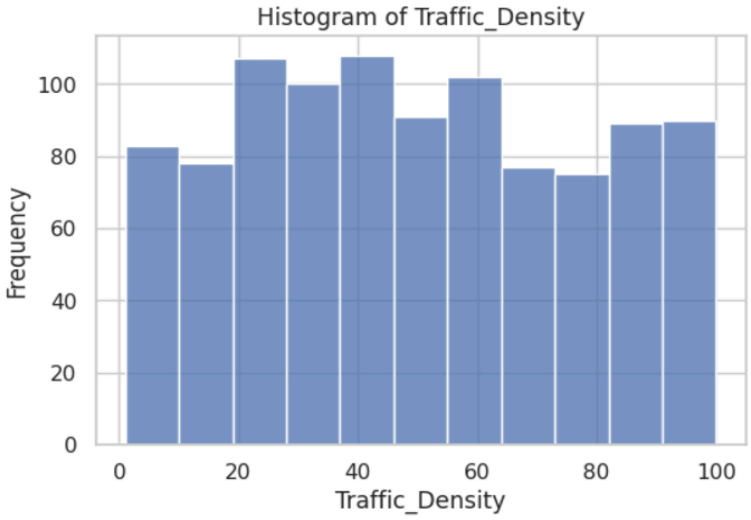

Fig 8.1 Histogram of Traffic Density

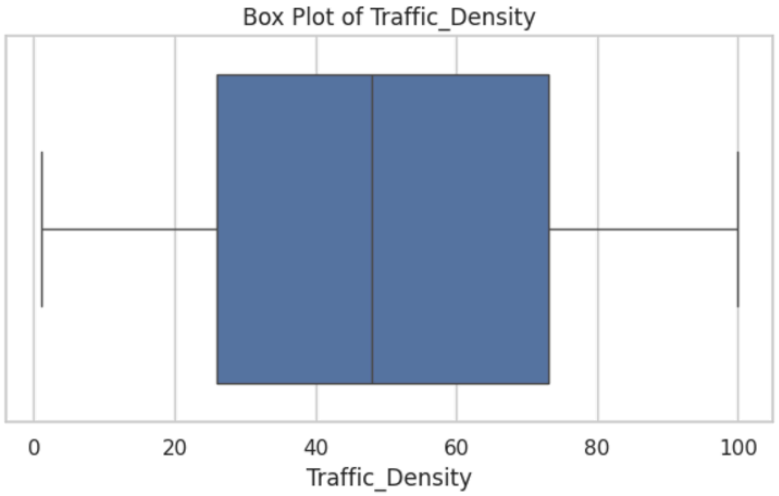

Fig 8.2 Box Plot of Traffic Density

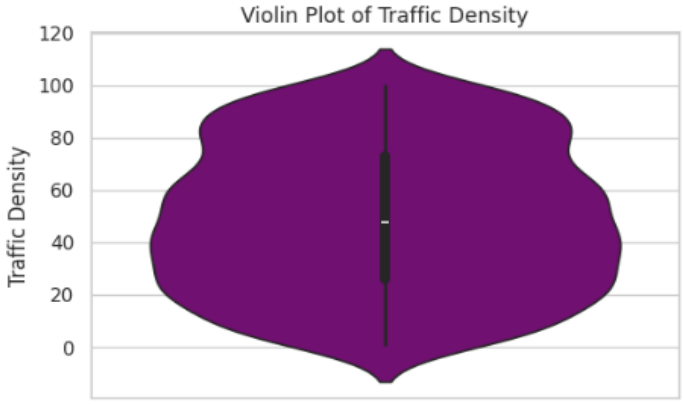

Fig 8.3 Violin plot of Traffic density

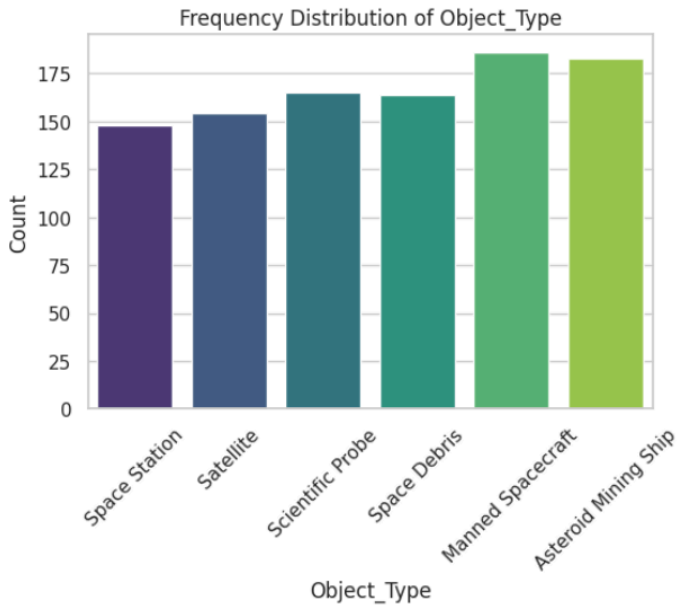

Fig 8.4 Frequency distribution of Object_Type

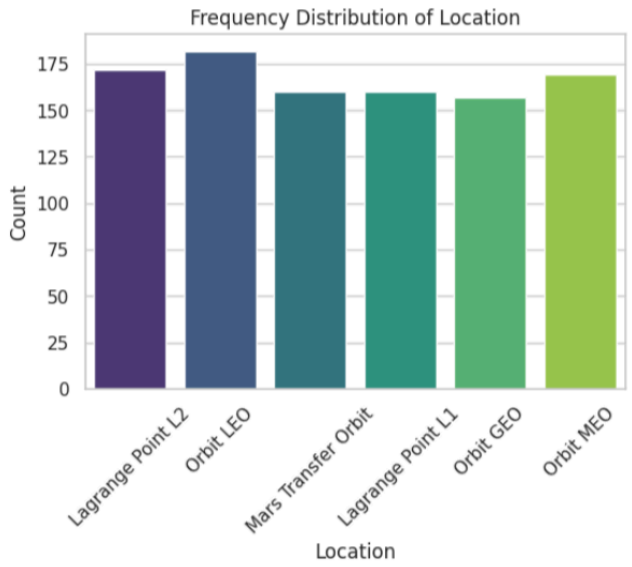

Fig 8.5 Frequency Distribution of Location

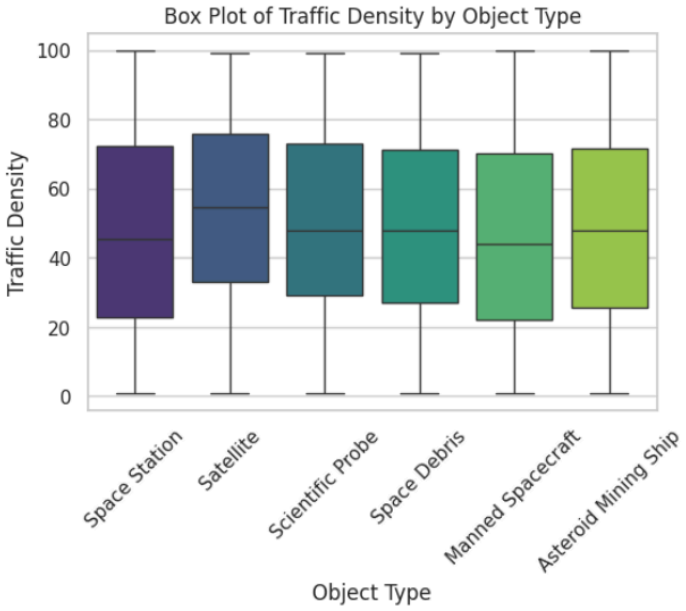

Fig 8.6 Box Plot of Traffic density by Object_Type

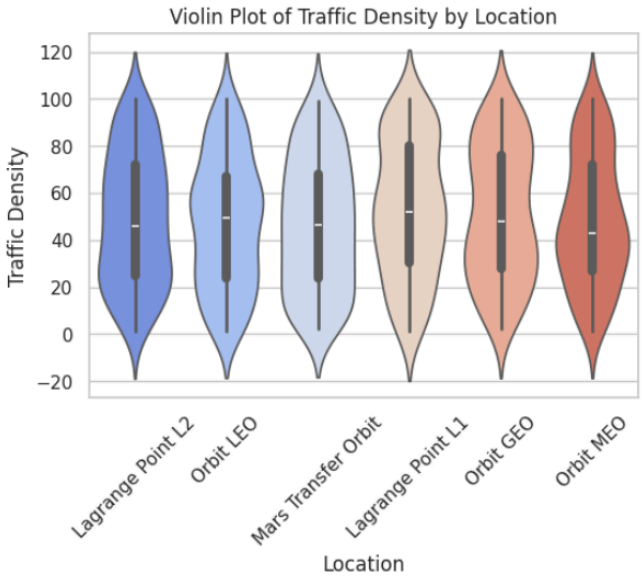

Fig 8.7 Box Plot of Traffic Density by Location

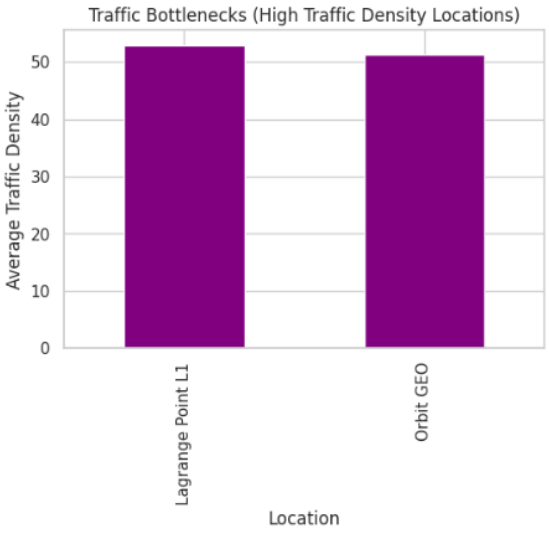

Fig 8.8 High Traffic Density Locations

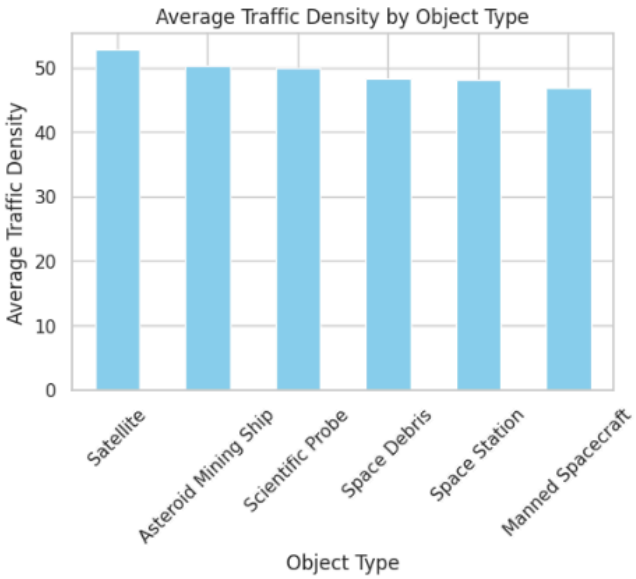

Fig 8.9 Average Traffic Density by Object_type

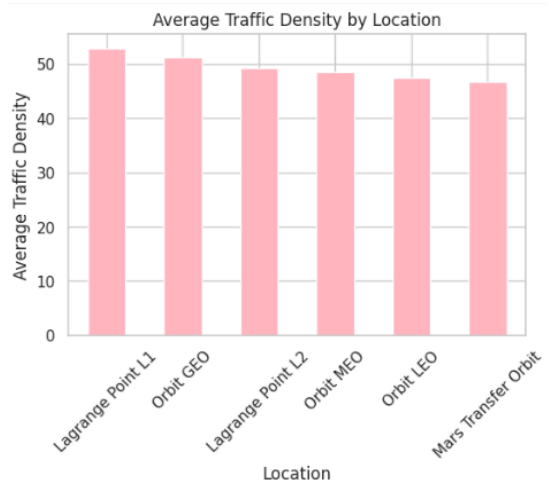

Fig 8.10 Average Traffic Density by location

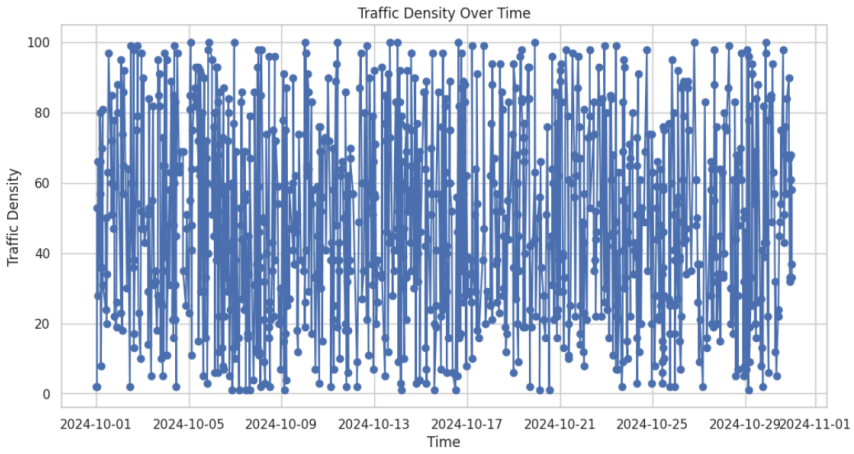

Fig 8.11 Traffic Density over time

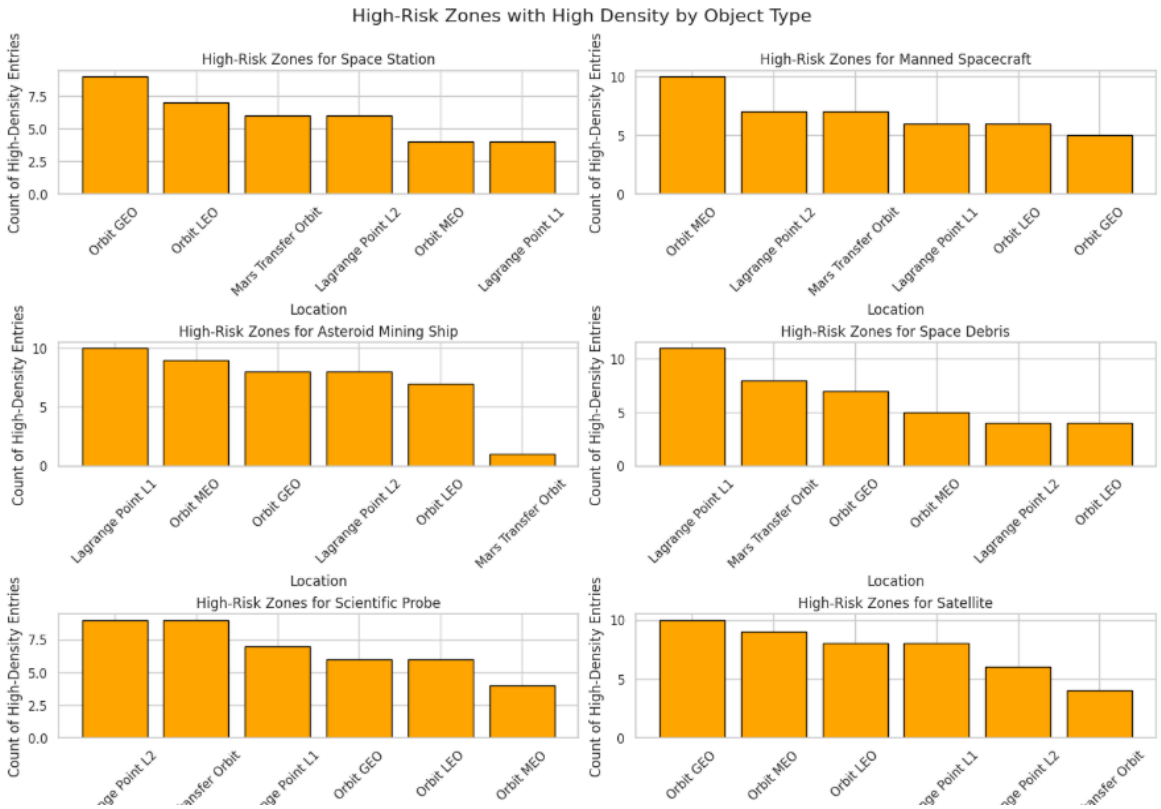

Fig 8.12 High Risk Zones with High Density by Object_type

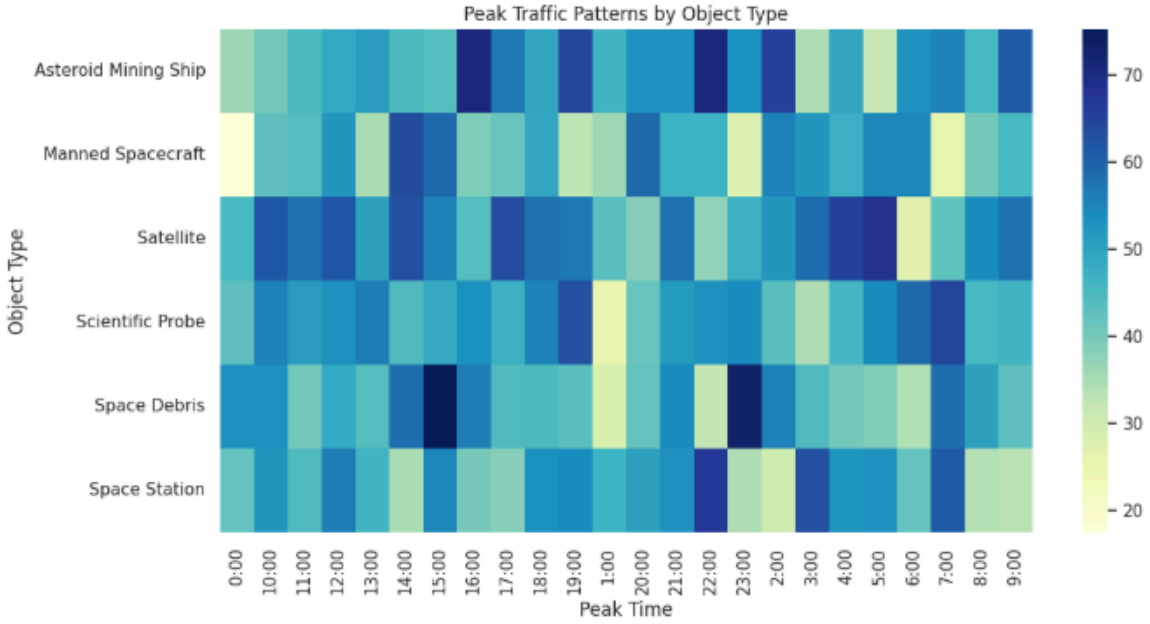

Fig 8.13 Peak Traffic Patterns by Object type

8. CONCLUSION

The first sprint successfully laid the foundation for predicting space traffic density. By gaining a comprehensive understanding of the dataset and exploring key traffic patterns, the project is well-positioned to transition into the model development phase in subsequent sprints.