**Problem Statement Document**

1. **Title**

**Analysis and Prediction of Space Traffic Density Using Regression Models**

**2. Objective**

The primary objective is to analyze the **Space Traffic Density Dataset**, perform data preprocessing and feature engineering, and build a **Linear Regression model** to predict traffic density accurately. This project involves cleaning the dataset, performing statistical analysis, creating features, and validating model robustness.

**3. Background**

With increasing space activities, understanding and predicting space traffic density is critical for efficient planning, collision avoidance, and operational safety. Accurate analysis of this dataset can help in better decision-making, resource optimization, and long-term planning for space traffic management.

**4. Problem Description**

The dataset consists of various attributes such as:

* **Traffic\_Density**: Numerical variable representing the space traffic density.
* **Object\_Type, Location, and Peak\_Time**: Categorical variables providing context about the traffic.

Key challenges:

1. The presence of **null values** and **outliers** that can affect the analysis.
2. **Complex relationships** between numerical and categorical variables.
3. **Model robustness** validation to ensure predictions are reliable.

To address these, the dataset will undergo preprocessing, analysis, and feature engineering, followed by a regression model to predict **Traffic\_Density**.

**5.Scope**

The following steps are covered in the analysis:

1. **Dataset Import and Understanding**: Import the Space Traffic Density dataset and explore its structure.
2. **Data Preprocessing**: Handle missing values, outliers, and categorical variables.
3. **Statistical Analysis**:
   * Perform univariate analysis for both numerical and categorical variables.
   * Conduct bivariate analysis to understand relationships.
   * Include time series analysis for trends.
4. **Feature Engineering**:
   * Create new features and encode categorical data.
   * Standardize features for model readiness.
5. **Model Development**:
   * Train a **Linear Regression model**.
   * Validate its robustness using metrics like R² and residual analysis.

**6. Dataset Overview**

The dataset contains the following attributes:

1. **Traffic\_Density**: A numerical target variable.
2. **Object\_Type**: Categories of objects in space traffic.
3. **Location**: Region where traffic is monitored.
4. **Peak\_Time**: Time periods with the highest traffic density.

**7. Methodology**

**Step 1**: Import and understand the dataset.  
**Step 2**: Clean the data:

* Impute null values using mean/median strategies.
* Handle outliers using IQR or capping methods.

**Step 3**: Conduct statistical analysis:

* Univariate analysis for traffic density and categorical variables.
* Bivariate analysis to understand relationships.

**Step 4**: Perform time series analysis to identify trends in traffic density.

**Step 5**: Feature Engineering:

* One-hot encoding and label encoding for categorical variables.
* Avoid the dummy variable trap.
* Standardize numerical features.

**Step 6**: Build and validate the linear regression model:

* Train the model using preprocessed data.
* Evaluate using metrics like R² and residual plots.

**8. Deliverables**

1. Cleaned and preprocessed dataset ready for modeling.
2. Statistical insights through visualizations and metrics.
3. A trained linear regression model with validated robustness.
4. Detailed report on the findings, challenges, and model performance.

**9. Challenges**

1. Identifying and handling missing data accurately.
2. Dealing with outliers to ensure data integrity.
3. Ensuring the regression model generalizes well to unseen data.

**10. Conclusion**

By following a structured approach, this project aims to accurately predict space traffic density while addressing key data quality issues and validating model robustness.