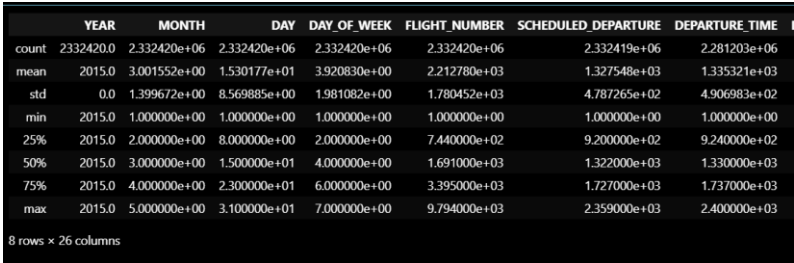


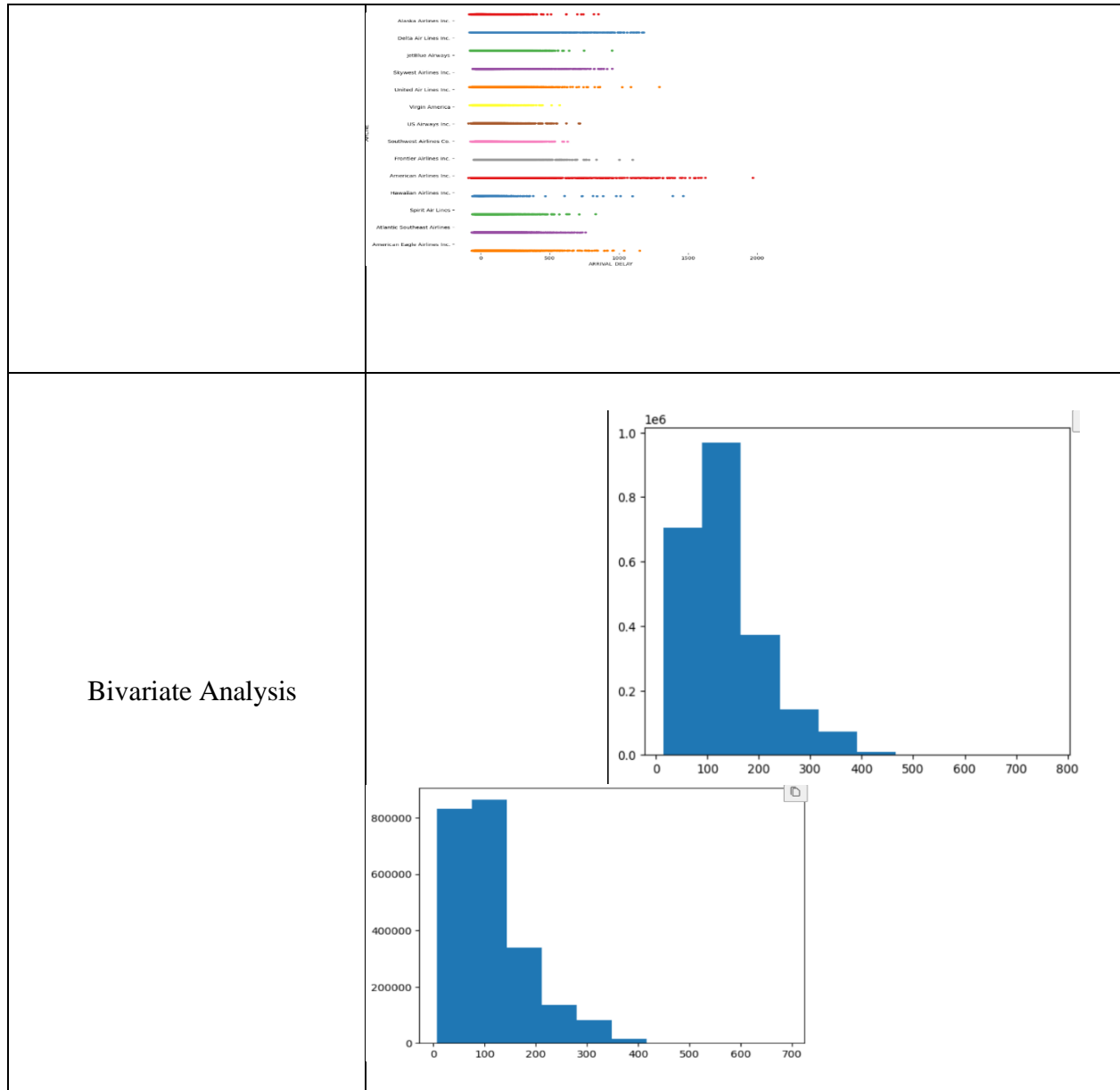
Data Collection and Preprocessing Phase

Date	23 September 2024
Team ID	LTVIP2024TMID24998
Project Title	Flight Delays Prediction using Machine Learning.
Maximum Marks	6 Marks

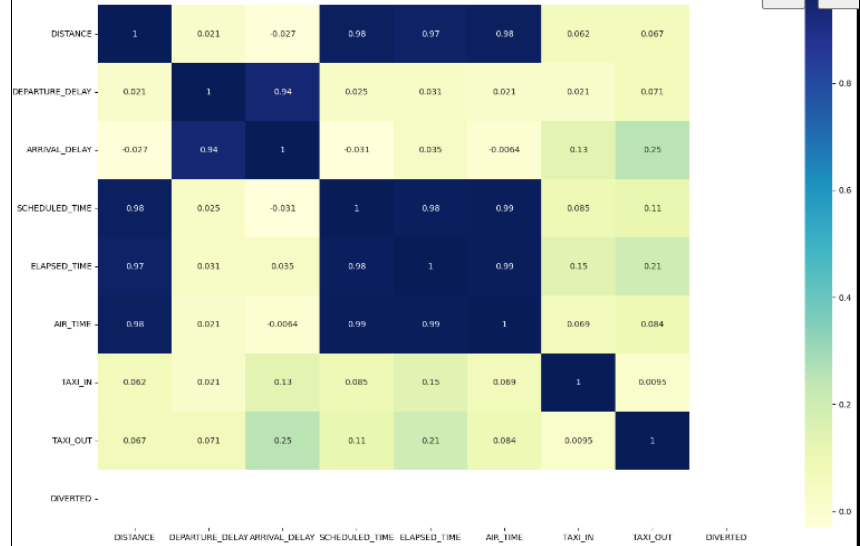
Data Exploration and Preprocessing Template

In the data exploration phase for Flight Delay Prediction, assess the dataset's structure, summarize key statistics, and visualize relationships and missing values. For preprocessing, handle missing data, engineer date features, and encode categorical variables to prepare the data for modeling.

Section	Description
Data Overview	 <p>8 rows x 26 columns</p>
Univariate Analysis	



Multivariate Analysis



Outliers and Anomalies

—

Data Preprocessing Code Screenshots

Loading Data

```
flightsinfo = pd.read_csv("flights.csv")
```

Handling Missing Data

```
flightsinfo_NULL = flightsinfo.isnull().sum()*100/flightsinfo.shape[0]
flightsinfo_NULL
```

```
flightsinfo1.isnull().sum()
```

Data Transformation

```
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
```

Feature Engineering

```
Flights1['AIRLINE'] = le.fit_transform(Flights1['AIRLINE'])
Flights1['ORIGIN_AIRPORT'] = le.fit_transform(Flights1['ORIGIN_AIRPORT'])
Flights1['DESTINATION_AIRPORT'] = le.fit_transform(Flights1['DESTINATION_AIRPORT'])
Flights1['Day'] = le.fit_transform(Flights1['Day'])
```

Save Processed Data	—
---------------------	---