Government Dataset

Analysis of Railway Stations: Connectivity, Layouts, and Infrastructure

INTRODUCTION:

This dataset provides structured information about **railway stations** in India, including details about their **zones, divisions, interchanges, station layouts, and distances**. The dataset consists of **54 railway stations**, with **12 attributes** that describe their connectivity and facilities.

->Key Features of the Dataset:

1.City – The location of the railway station.

1.Zone – The railway zone under which the station operates.

2.Division – The railway division that manages the station.

3.Connection – The major railway line(s) the station is part of.

4.Interchange – Other railway lines or metro networks connected to the station.

5.Station & Station Code – Name and code of the station.

6.Distance in Kms – The station's distance from a reference point.

7.Layout – The type of railway station layout (e.g., **At Grade**).

8.Parking Availability – Indicates whether parking is available.

INTRODUCTION:

Purpose of the Dataset:

- **Railway Planning & Analysis**: Helps identify important **interchange hubs** and **station layouts**.
- **Urban Transport Development**: Useful for studying railway connectivity and infrastructure availability.
- **Operational Insights**: Zones and divisions allow authorities to optimize routes and passenger convenience.

Overview:

	City	Zone	Division	Connection	Interchange	Station	Station Code	Distance in Kms	Layout	Parking	Contract Available Space Avaiable	No Contract Sapce Not Avaiable
0	Chennai	5	60	South Line	North Line; West Line; MRTS Line	Chennai Beach RS	MSB	0.00	At Grade	Yes	No	No
1	Chennai	5	60	South Line	MRTS Line	Chennai Fort RS	MSF	1.80	At Grade	No	No	Yes
2	Chennai	5	58	South Line	North Line; West Line; MRTS Line; CMRL Blue; C...	Chennai Park RS; Chennai Central	MPK	3.07	At Grade	Yes	No	No
3	Chennai	5	61	South Line	CMRL Green Line	Chennai Egmore RS; Rail Terminus	MS	4.32	At Grade	Yes	No	No
4	Chennai	8	104	South Line	No	Chetpat RS	MSC	6.56	At Grade	Yes	No	No

Column Description:

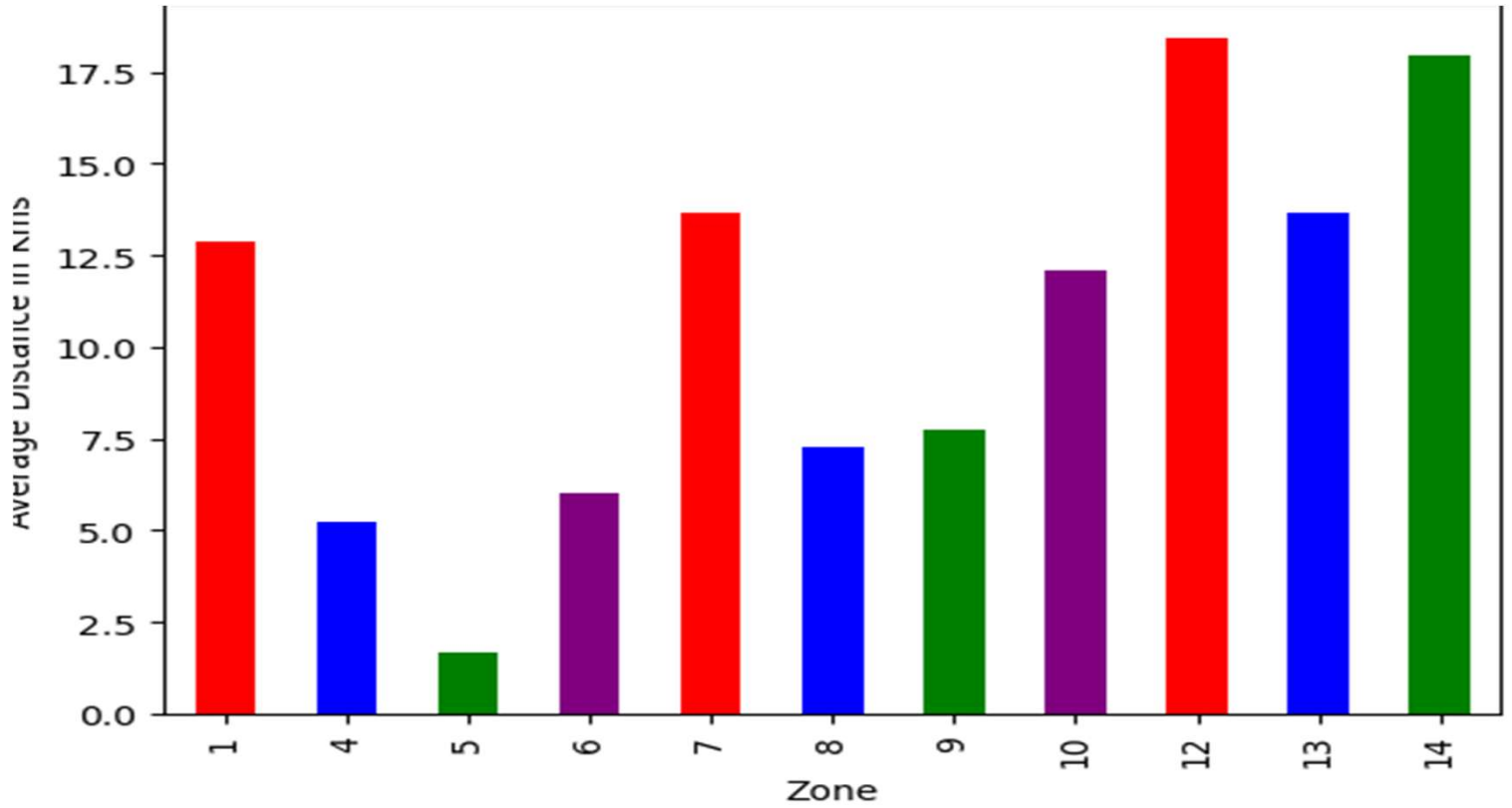
Column Name	Description
City	The name of the city where the railway station is located.
Zone	The railway zone under which the station operates (e.g., Southern Railway, Northern Railway).
Division	The railway division responsible for managing the station.
Connection	The major railway line(s) to which the station belongs.
Interchange	Other railway lines or metro networks connected to the station. If multiple, they are separated by a semicolon (`;`).
Station	The name of the railway station.
Station Code	The official code assigned to the railway station.
Distance in Kms	Distance of the station (likely from a reference point, such as a main station or starting point of a route).
Layout	The structural design of the railway station (e.g., **At Grade**, Elevated, Underground).
Parking Contract Available	Indicates whether parking is available under a contract (`Yes` or `No`).
Space Available No Contract	Indicates whether parking space is available without a contract (`Yes` or `No`).
Space Not Available	Indicates whether there is no parking space available at the station (`Yes` or `No`).

Data Cleaning:

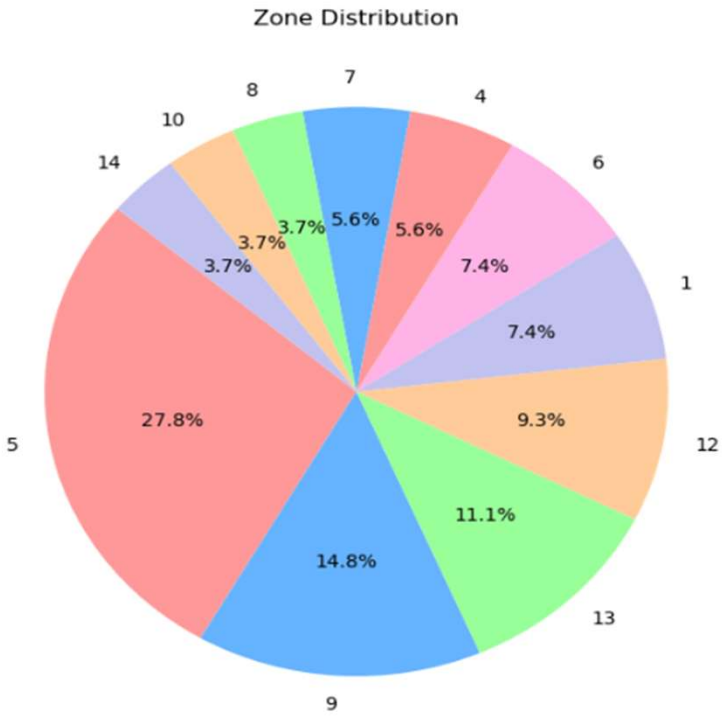
```
for col in ["Parking Contract Available", "Space Available No Contract", "Space Not Available", "Layout", "Connection", "Interchange"]:
    Railway[col] = Railway[col].fillna("Unknown").str.title()
Railway.drop_duplicates(inplace=True)
Railway["Distance in Kms"] = pd.to_numeric(Railway["Distance in Kms"], errors="coerce")
Railway.info(), Railway.isnull().sum()
```

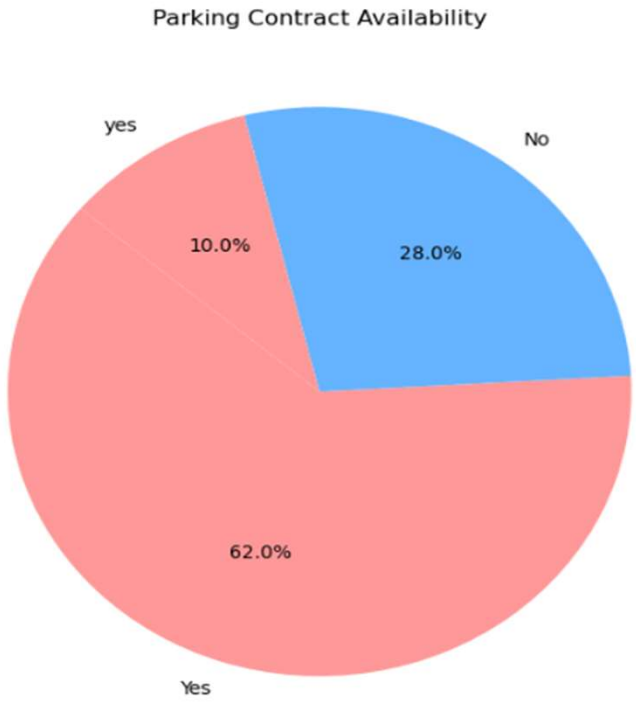
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54 entries, 0 to 53
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   City                            54 non-null     object 
 1   Zone                            54 non-null     int64  
 2   Division                        54 non-null     int64  
 3   Connection                      54 non-null     object 
 4   Interchange                     54 non-null     object 
 5   Station                         54 non-null     object 
 6   Station Code                    54 non-null     object 
 7   Distance in Kms                 54 non-null     float64
 8   Layout                          54 non-null     object 
 9   Parking Contract Available      54 non-null     object 
 10  Space Available No Contract     54 non-null     object 
 11  Space Not Available             54 non-null     object 
 12  Interchange Count               54 non-null     int64  
 13  Station Sequence                54 non-null     int64  
dtypes: float64(1), int64(4), object(9)
memory usage: 6.0+ KB
(None,
 City                            0
 Zone                            0
 Division                        0
 Connection                      0
 Interchange                     0
 Station                         0
 Station Code                    0
 Distance in Kms                 0
 Layout                          0
 Parking Contract Available      0
 Space Available No Contract     0
 Space Not Available             0
 Interchange Count               0
 Station Sequence                0
 dtype: int64)
```

Average Distance by zone:

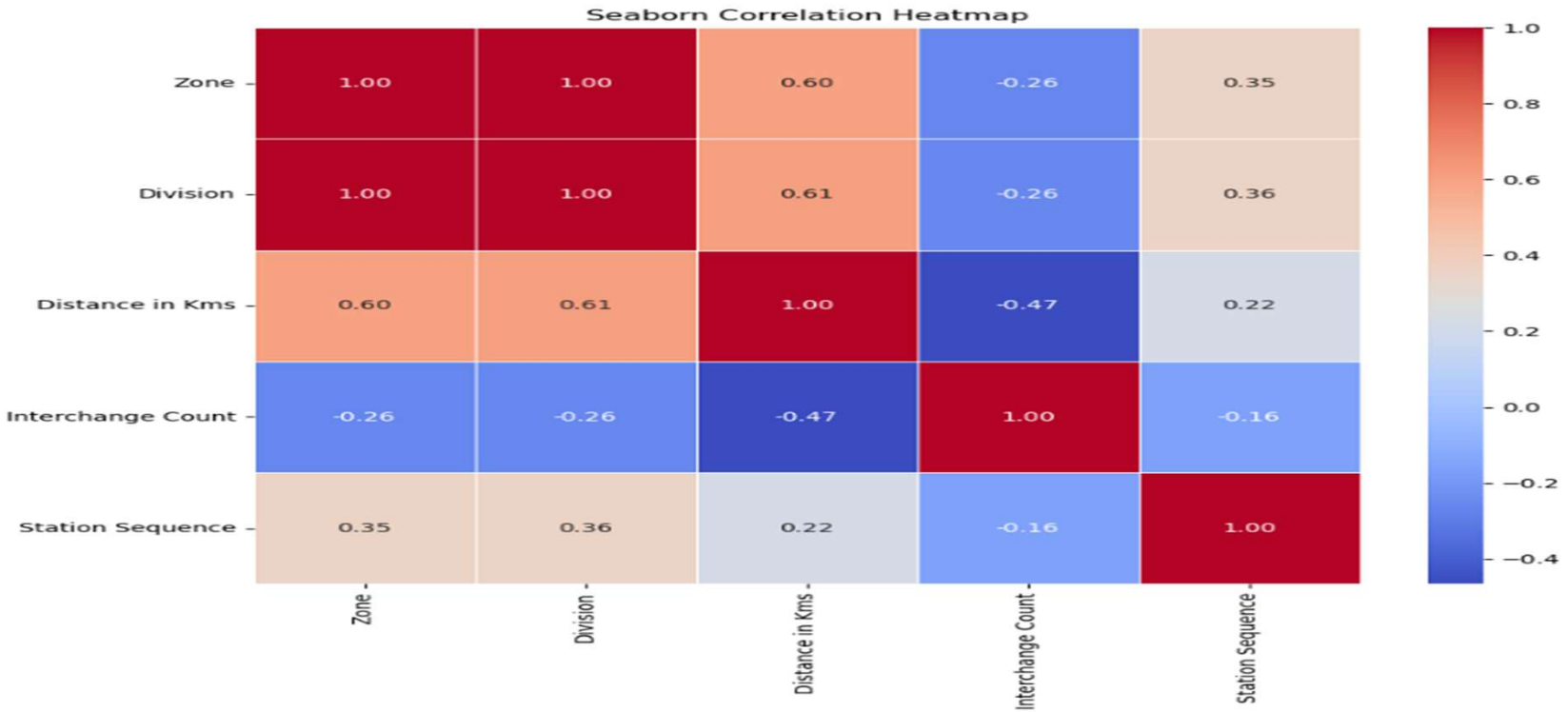


Parking & zone:

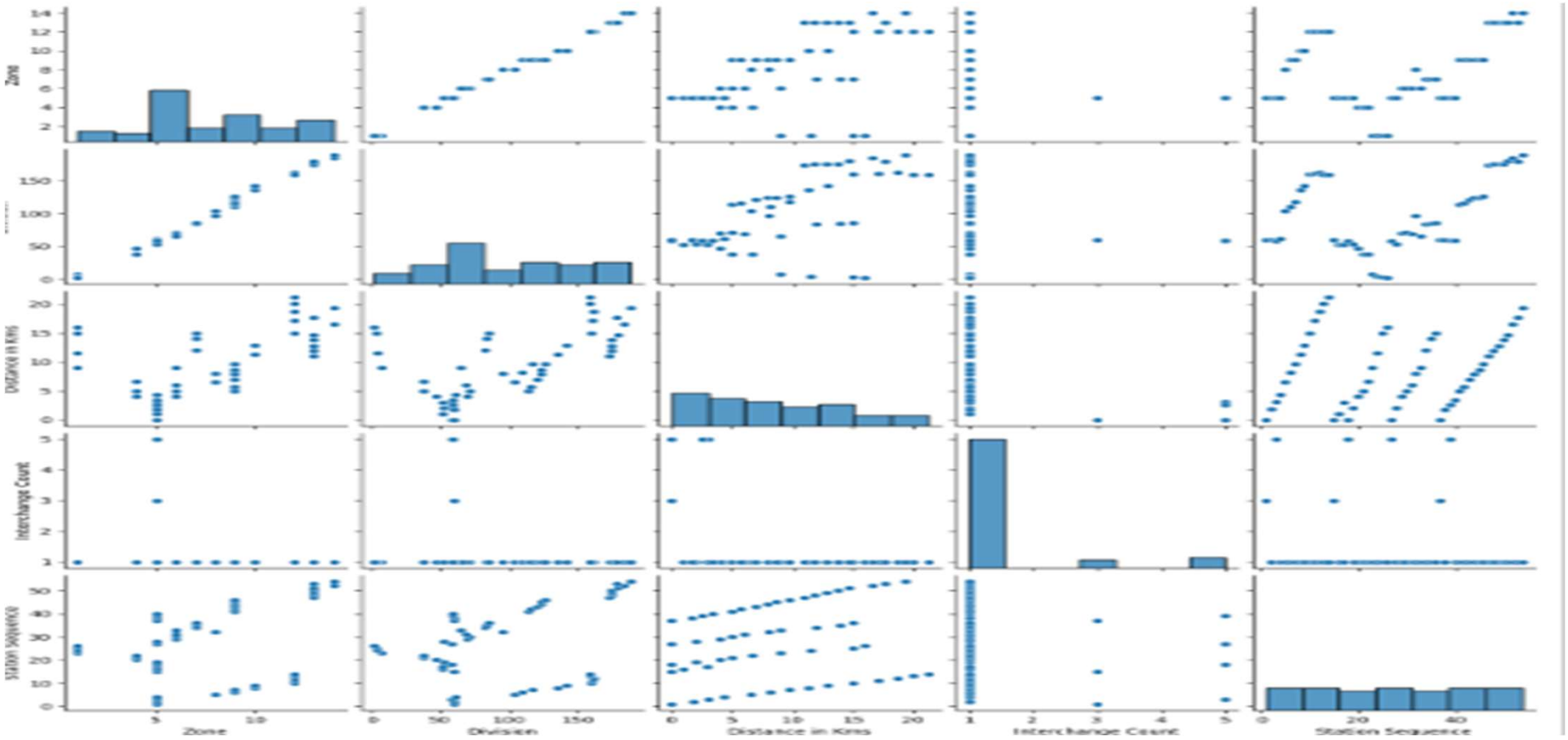




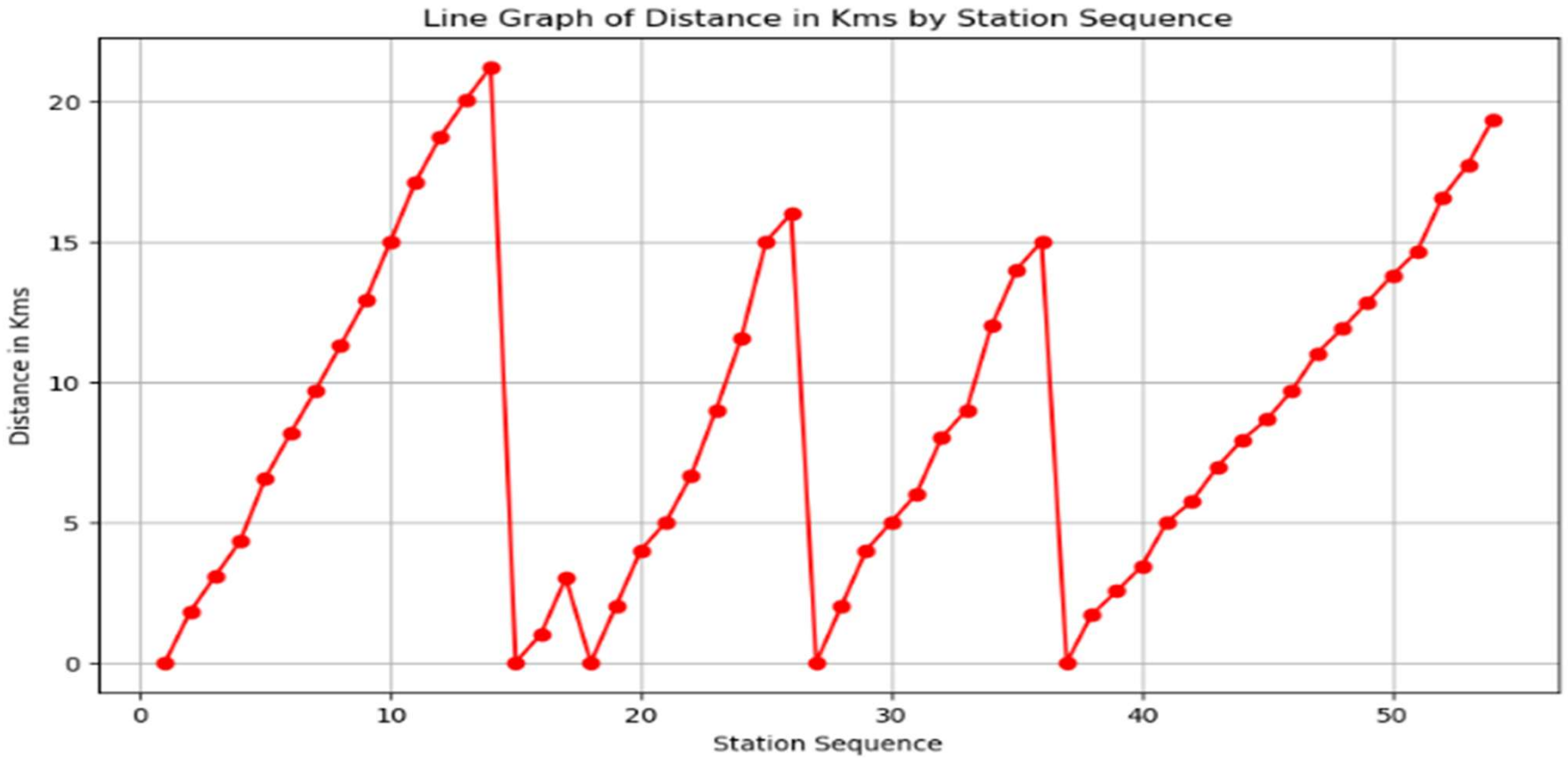
Heatmap:


Seaborn Correlation Heatmap
Zone
Division
Distance in Kms
Interchange Count
Station Sequence
1.00 1.00 0.60 -0.26 0.35
1.00 1.00 0.61 -0.26 0.36
0.60 0.61 1.00 -0.47 0.22
-0.26 -0.26 -0.47 1.00 -0.16
0.35 0.36 0.22 -0.16 1.00
Zone
Division
Distance in Kms
Interchange Count
Station Sequence
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

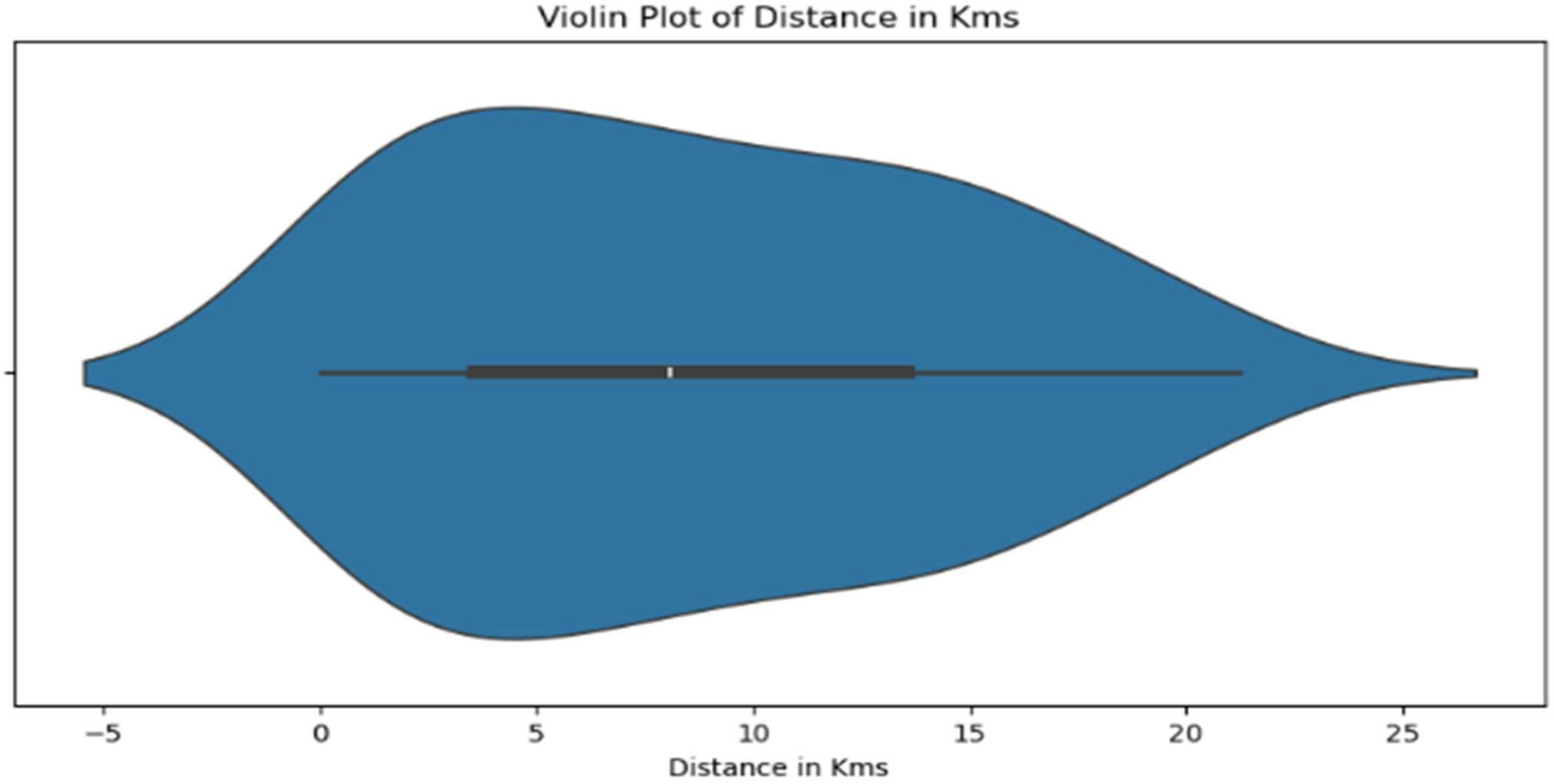
Pairplot Railway Dataset:



Line Graph:


Line Graph of Distance in Kms by Station Sequence
Distance in Kms
Station Sequence
0
5
10
15
20
0
10
20
30
40
50

Vilon Plot:



Railway Data Analysis Report:

•Missing values in categorical columns like "Parking Contract Available," "Space Available No Contract,“ and "Layout" were filled with "Unknown." There are no missing values remaining after cleaning.

•Column names were corrected, and text values in categorical columns were standardized to title case for consistency.Duplicate records were removed from the dataset.

•The "Distance in Kms" column was converted to numeric format, ensuring there are no incorrect or non-numeric values.

•The dataset shows a wide range of values in "Distance in Kms," indicating some stations are significantly farther than others. The most common station layouts were identified through the "Layout" column. The dataset provides information about railway stations, parking availability, and connectivity, which can be used for further analysis.

•The pairplot visualization suggests possible correlations between distance, connectivity, and station layout. It also shows clustering of stations based on zone or interchange type, indicating potential relationships among variables.

THANK YOU