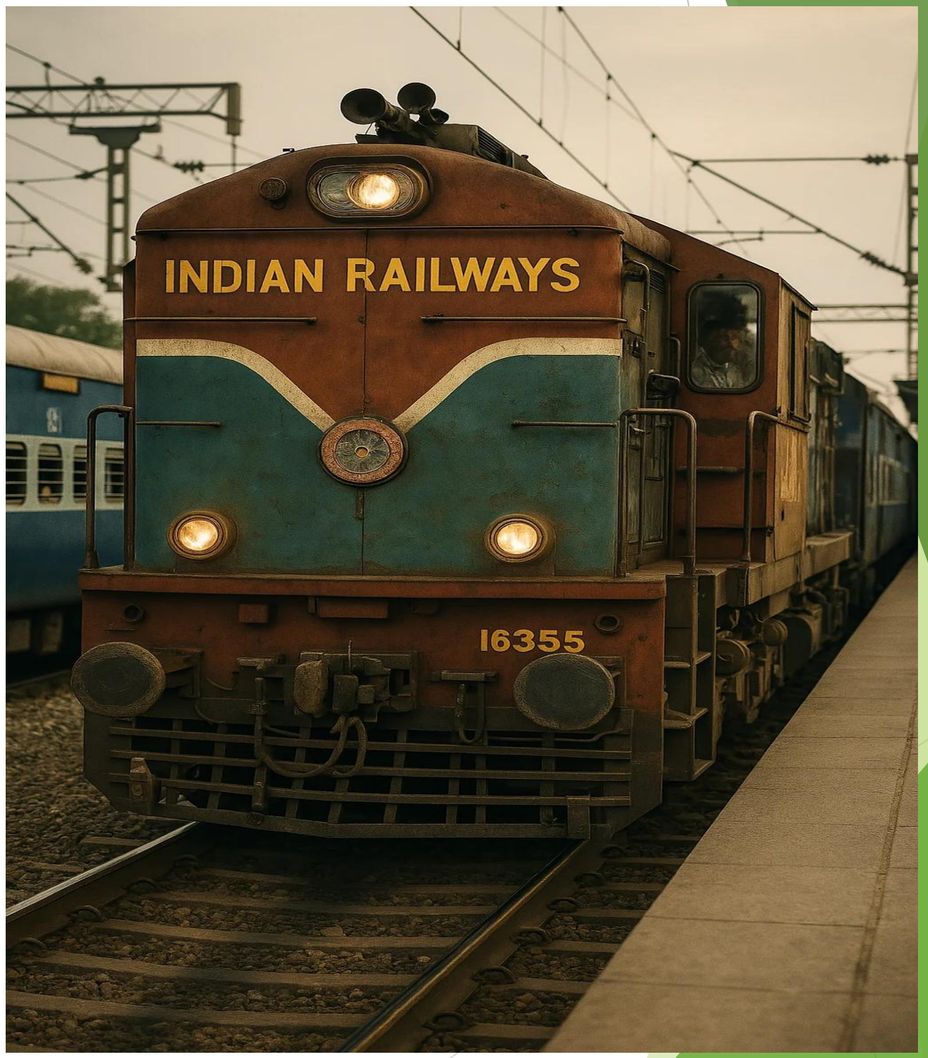
INDIAN RAILWAYS

By- SHASWAT SINGH BT23CSA030
HITANSH JAINBT23CSA034

Introduction

This report provides a comprehensive analysis of Indian Railways' passenger data from the fiscal year 2003–04 to 2023–24. The goal is to visualize trends in passenger count, identify growth patterns, and understand operational scale. Web scraping was used to extract historical data from reliable online sources.



Problem Description

- This project focuses on analyzing multiple datasets related to Indian Railways to gain insights into its operations, performance, and resource management.
- Indian Railways is one of the world's largest rail networks, carrying millions of passengers daily and operating thousands of trains across the country.
- The goal is to uncover patterns and trends that can support strategic decision-making and help optimize railway infrastructure and services.
- The project will also benchmark Indian Railways against global railway systems to evaluate performance in areas like efficiency, service quality, and coverage.

TOPICS OF ANALYSIS

RAILWAY

- Geographic Distribution of Railway Infrastructure
- Temporal Patterns and Operational Efficiency
- Train Performance and Characteristics
- Station Popularity and Connectivity
- Resource Allocation and Comparative Analysis

Dataset Description

This presentation analyzes multiple datasets to understand Indian Railways' operations, passenger flow, and resource allocation. We use train schedules, station locations, train details, passenger counts, employee numbers, budget allocations, and global railway line lengths.

Data Exploration & Preprocessing:

* Cleaned data, handled missing values, and standardized data types.
* Created GeoDataFrames for spatial analysis (station locations, train routes).
* Merged datasets for a comprehensive overview.

Key Findings (Visualizations):

Geographic Distribution: Maps show station density, train routes, and state-wise network distribution. Compared infrastructure to population density.

Temporal Trends: Time-series plots show passenger counts, budget allocations, and employee trends over time, revealing seasonal variations and growth patterns. Train arrival/departure analysis highlights operational bottlenecks.

Train Performance: Visualizations compare train speeds, distances, and durations; showing correlations with passenger numbers and class availability.

Station Popularity: Identified key hubs and connectivity patterns, using visualizations.

Resource Allocation: Analyzed budget allocation trends and relationships with employee numbers.

Global Comparison:Compared India's railway network (line length) to other countries.

Exporting Cleaned Data

1.Data Import & Initial View
- •Loaded dataset using pandas.
- •Viewed initial rows to understand structure (head()).

2.Basic Info & Shape
- •Checked dataset shape (rows, columns).
- •Used info() and describe() to inspect datatypes and basic stats.

3.Missing Value Analysis
- •Identified columns with missing values using isnull().sum().

4.Unique Values & Duplicates
- •Counted unique values per column.
- •Checked for and dropped duplicate rows.

5.Categorical & Numerical Columns
- •Separated categorical vs numerical columns for better analysis.

6.Univariate Analysis
- •Plotted bar plots, histograms for individual features like zone, year, etc.

7.Bivariate Analysis
- •Explored relationships between features (e.g., Year vs Passengers) using line plots and group-by.

8.Trend Analysis
- •Analyzed how passenger numbers changed year-wise and zone-wise.

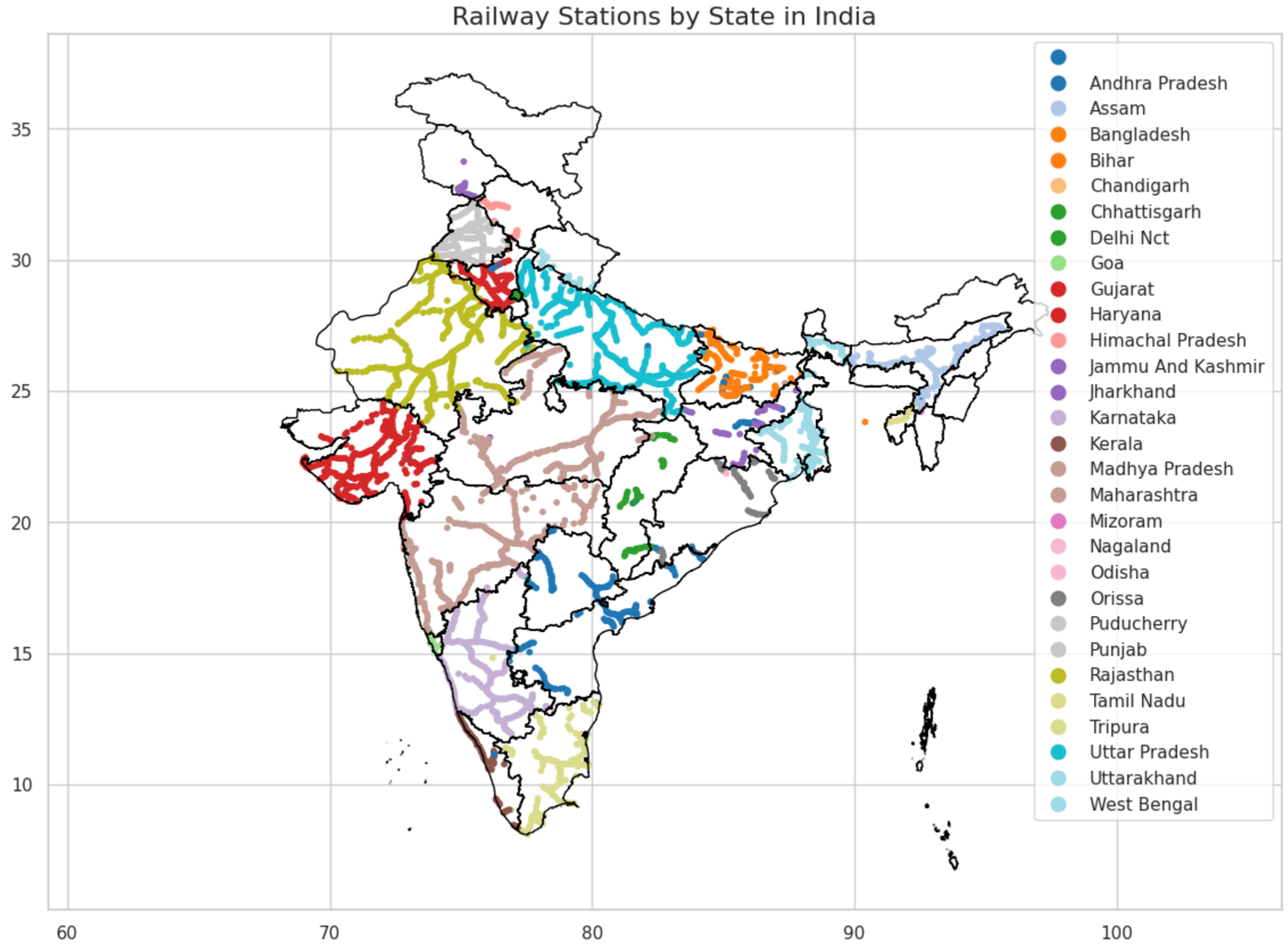
9.Outlier Detection
- •Looked at numerical columns using boxplots to find outliers.

10.Correlation Analysis

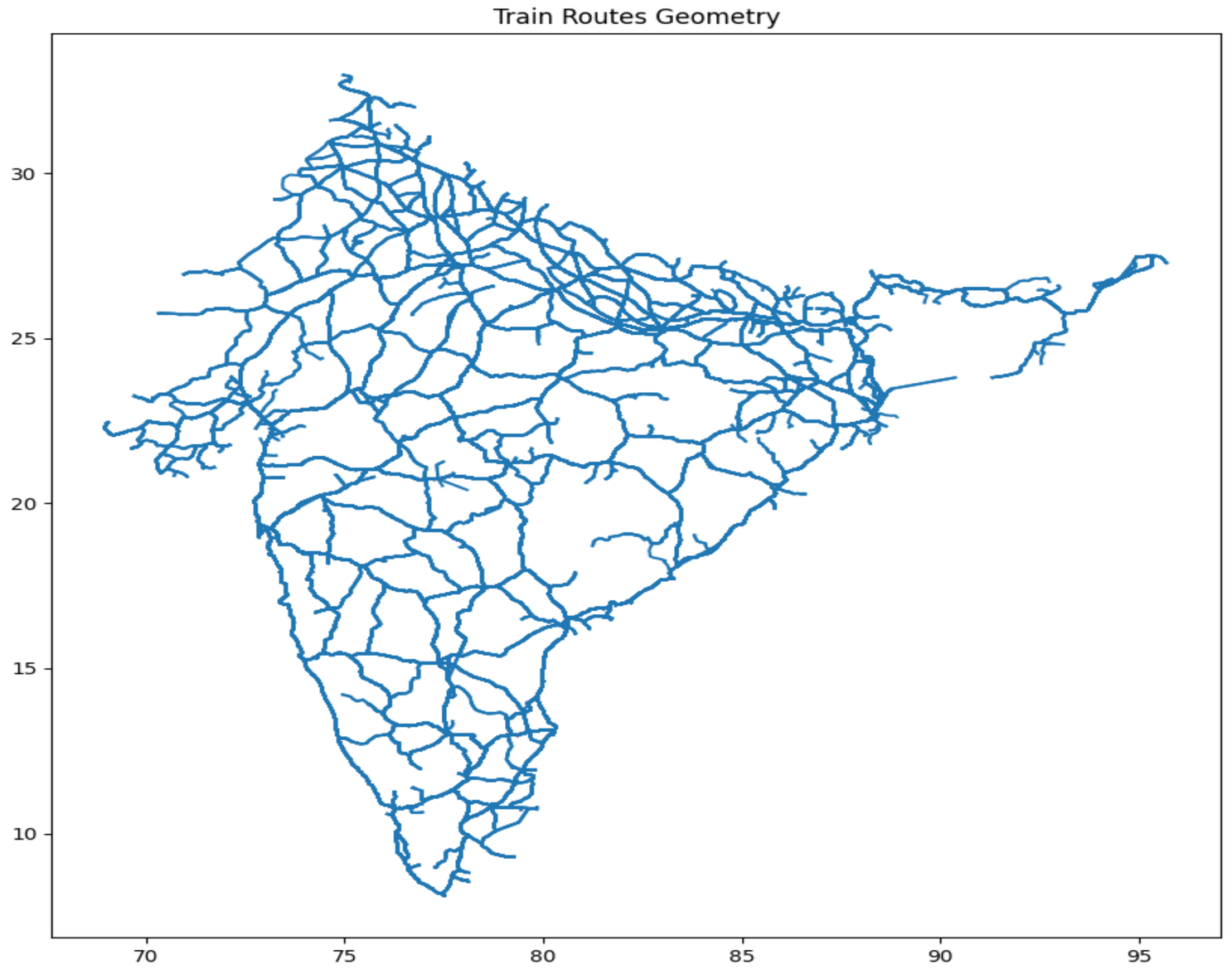
•Used heatmap to check correlation between numerical features (if any).

TOP 21 QUESTIONS

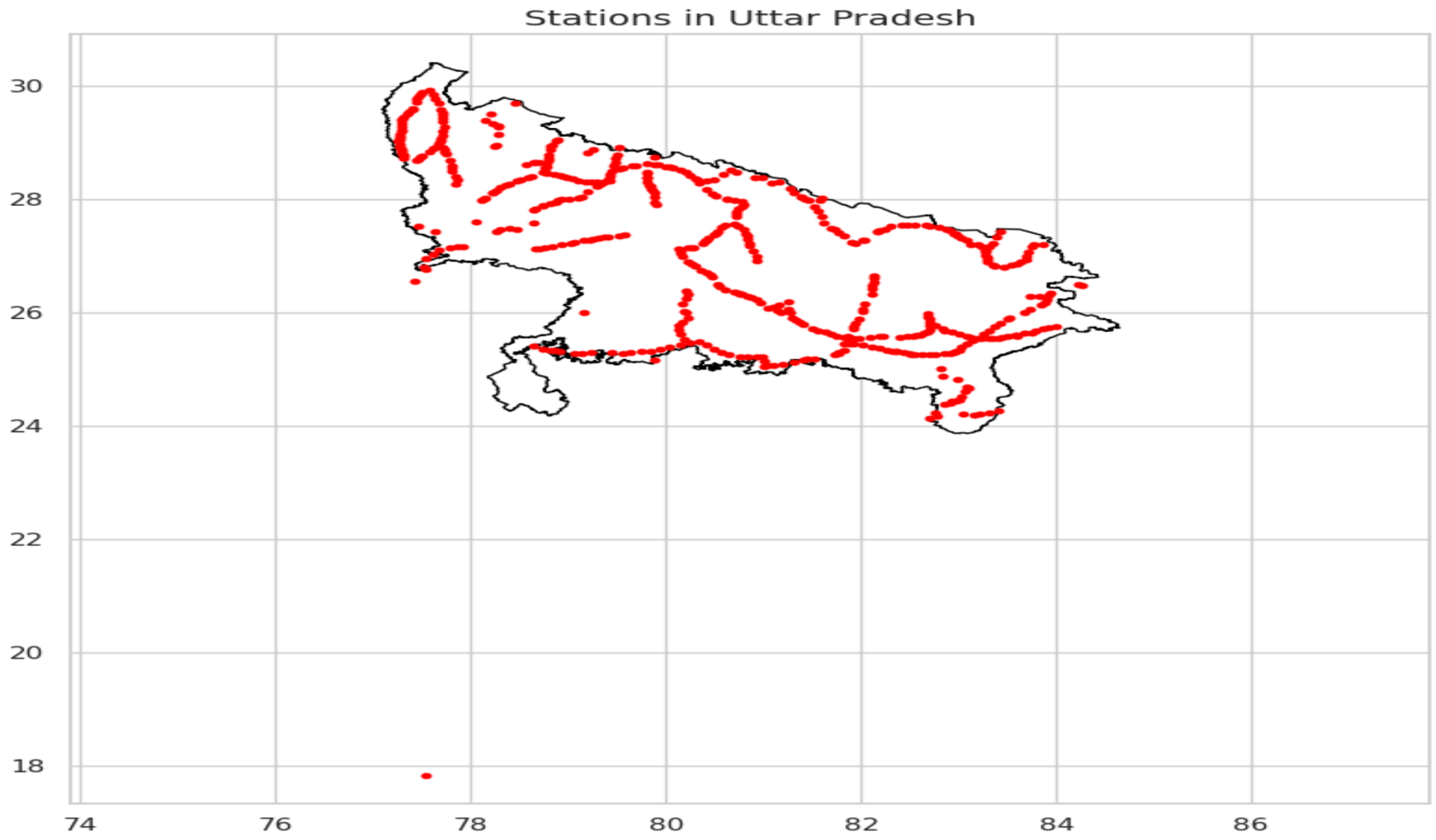
Q1. Plot stations location in India?



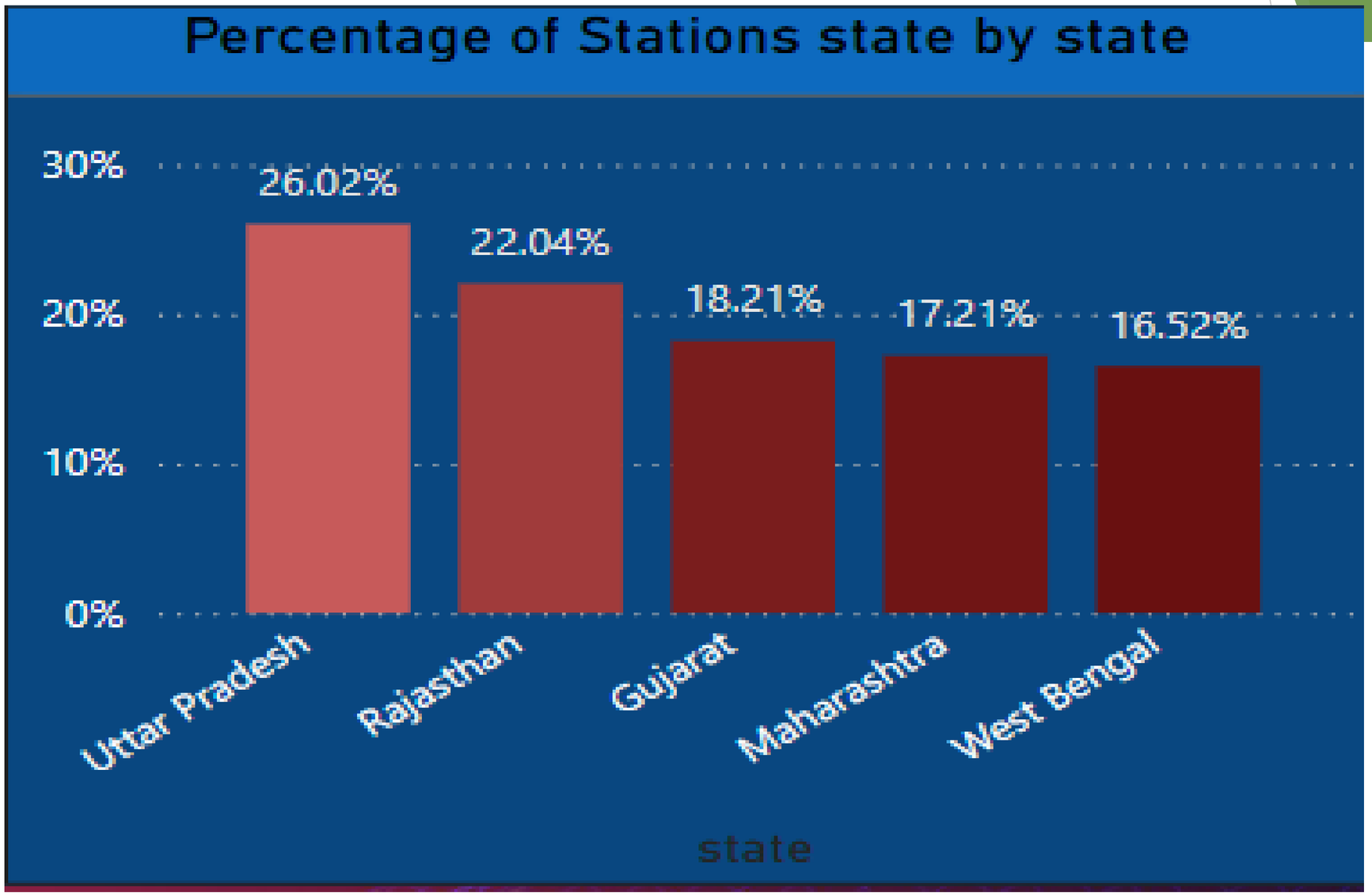
Q2. What is Geometry of Train Routes in India?



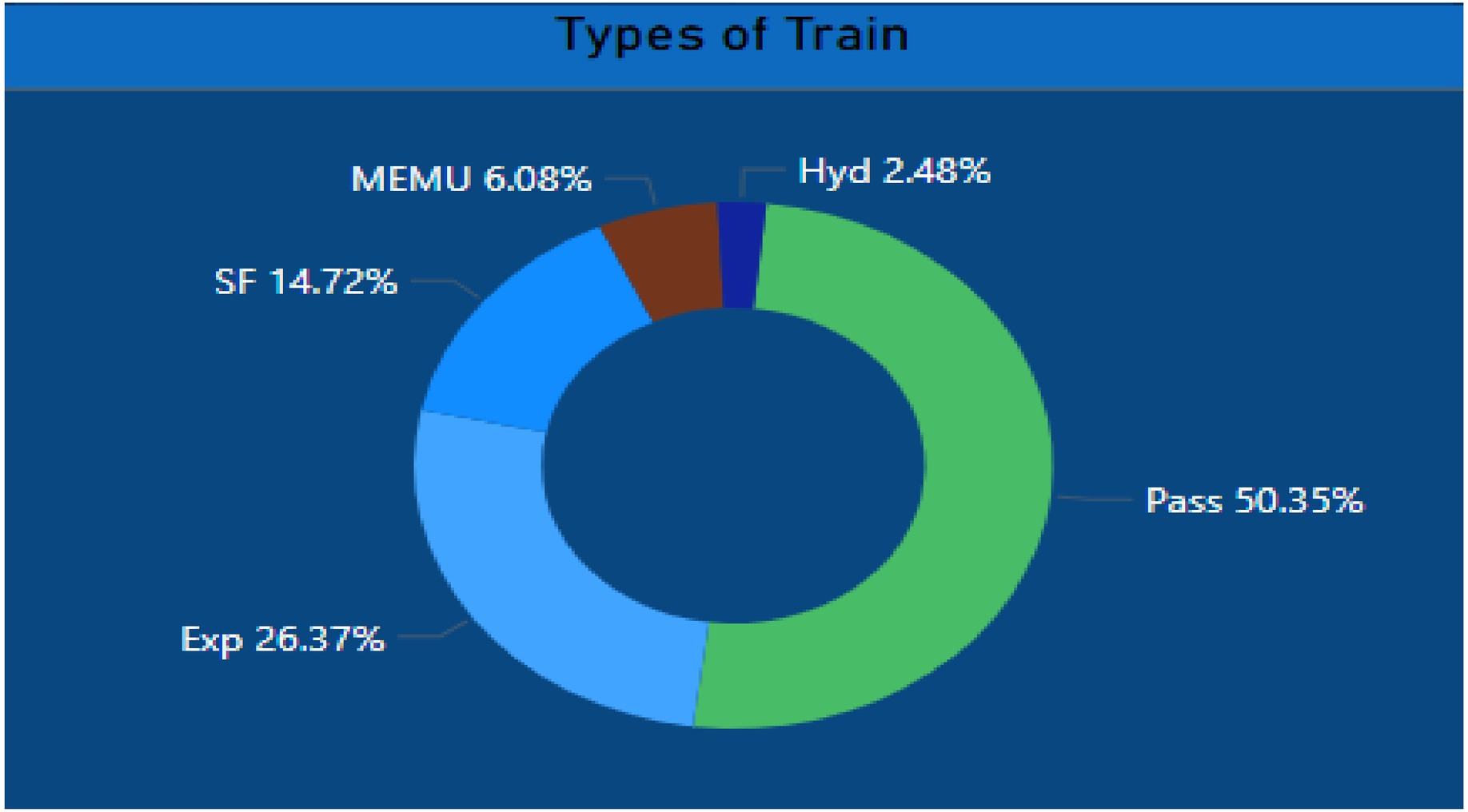
Q3. Map of State having highest number of stations



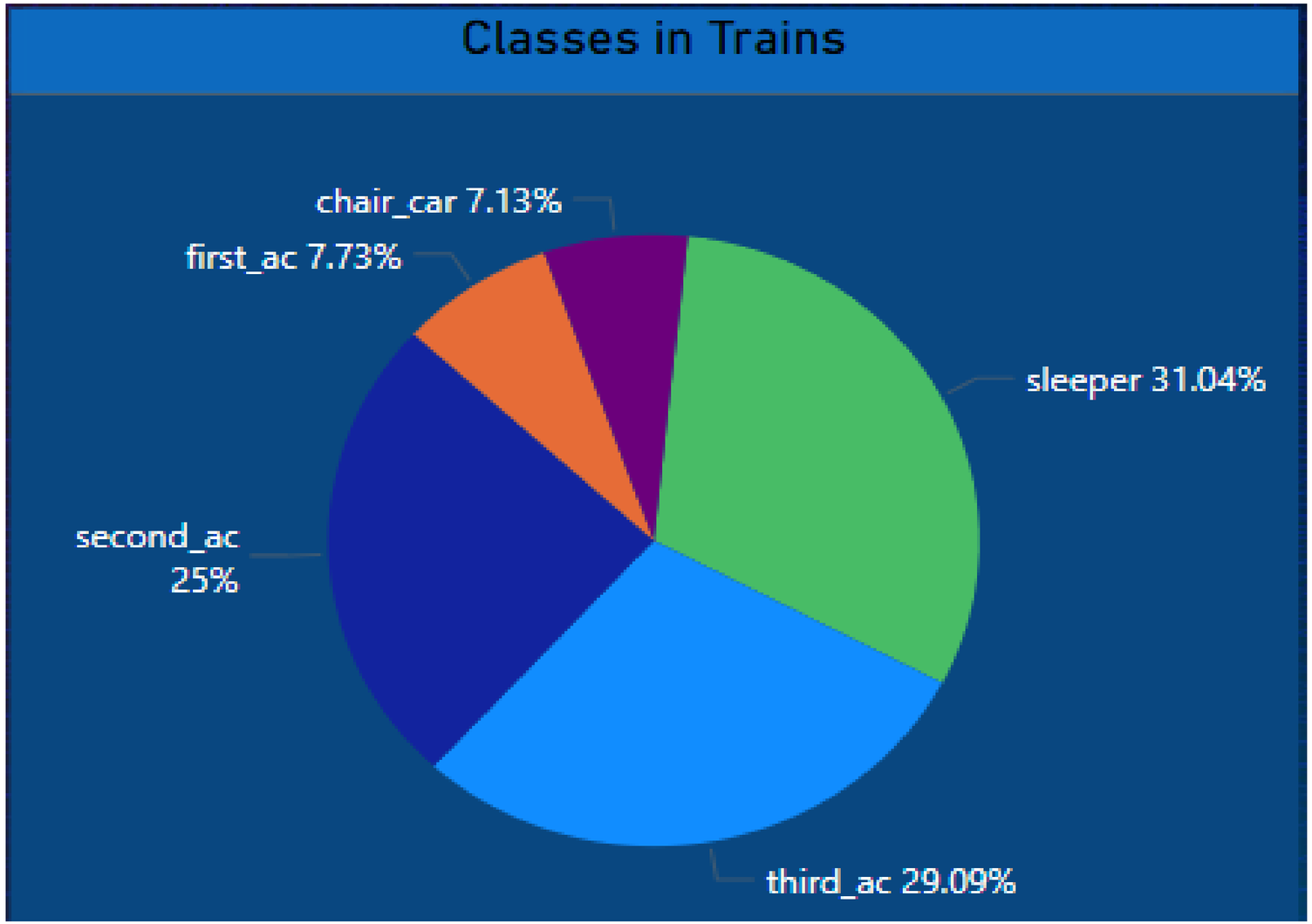
Q4. Which states have the most stations?



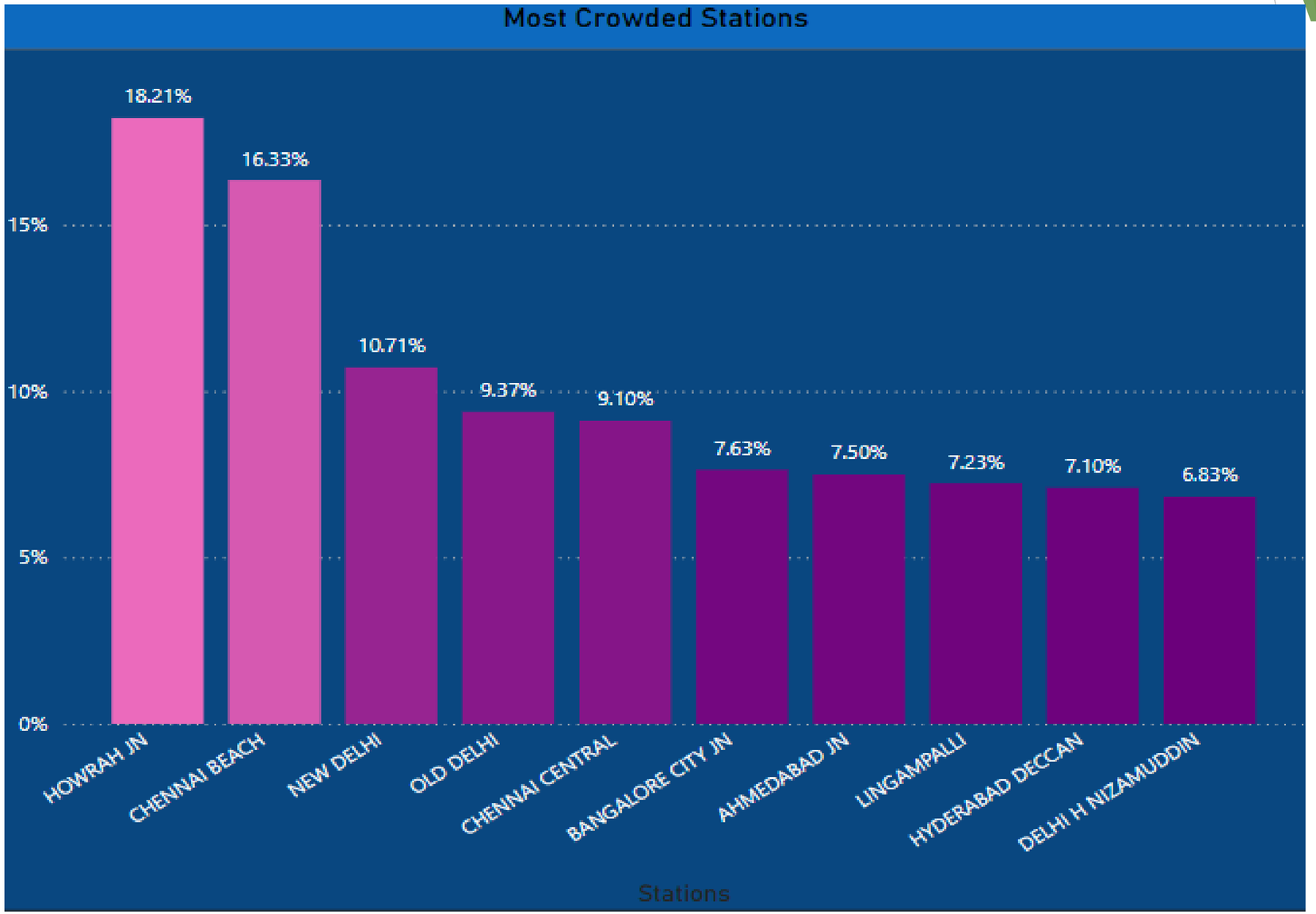
Q5. which are most common classes in Train?



Q6. which are most common classes in Train?

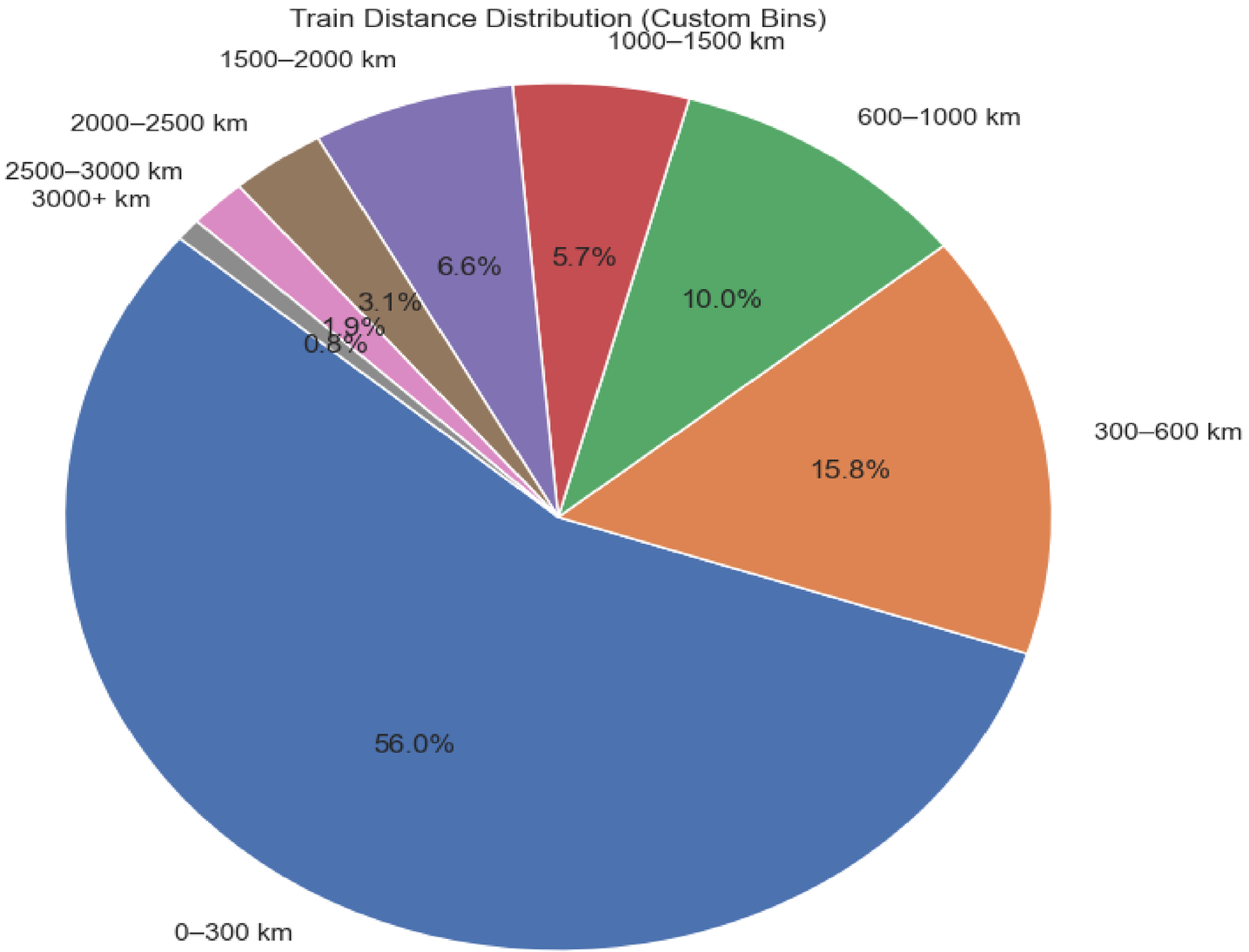


Q7. Which are 10 most Crowded stations?

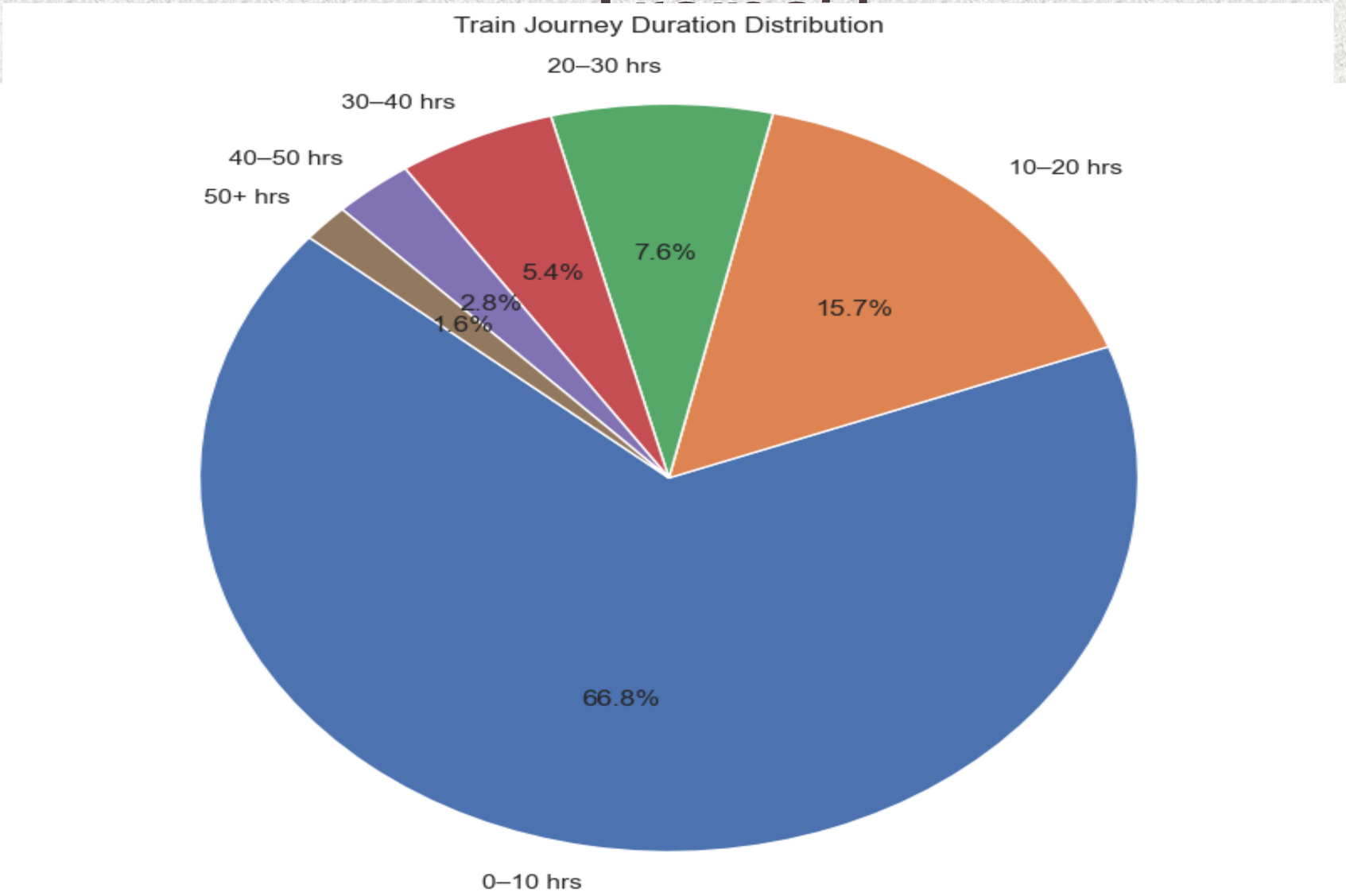


Q8. What are the number ofpeople use train in an year(bn)?

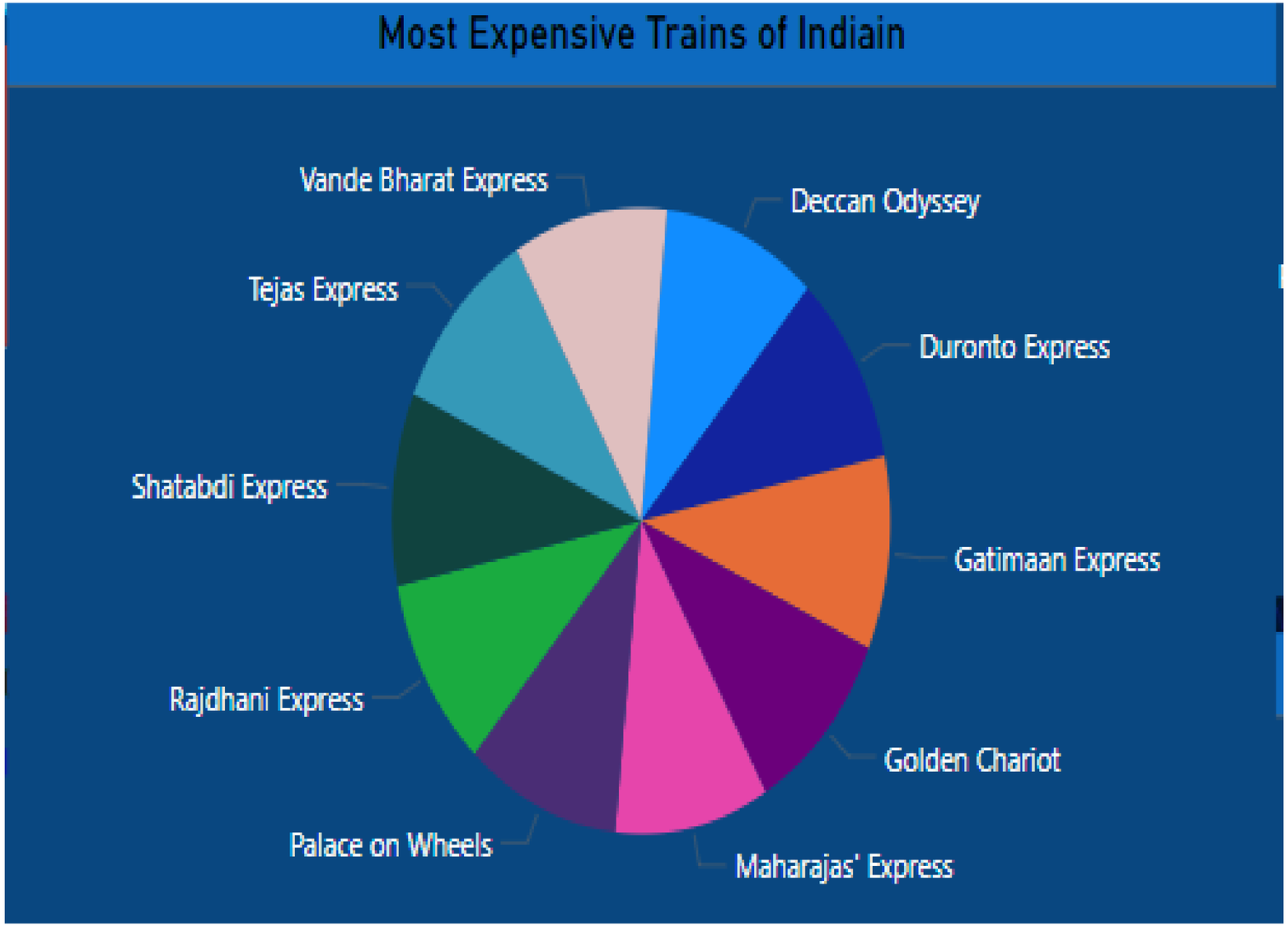
Q9. What is distance distribution of Indian Trains?



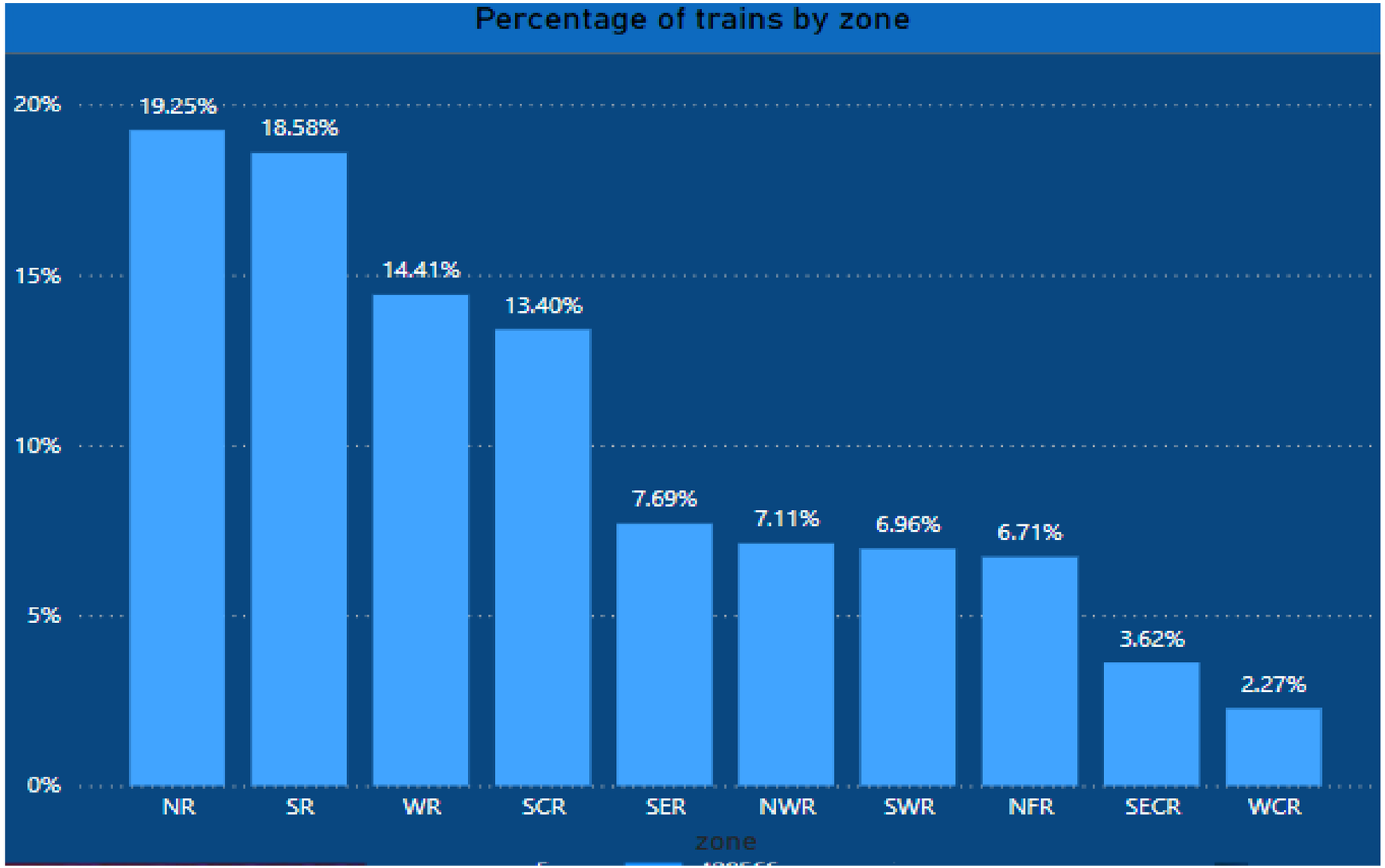
Q10. What is duration distribution of Indian Trains?



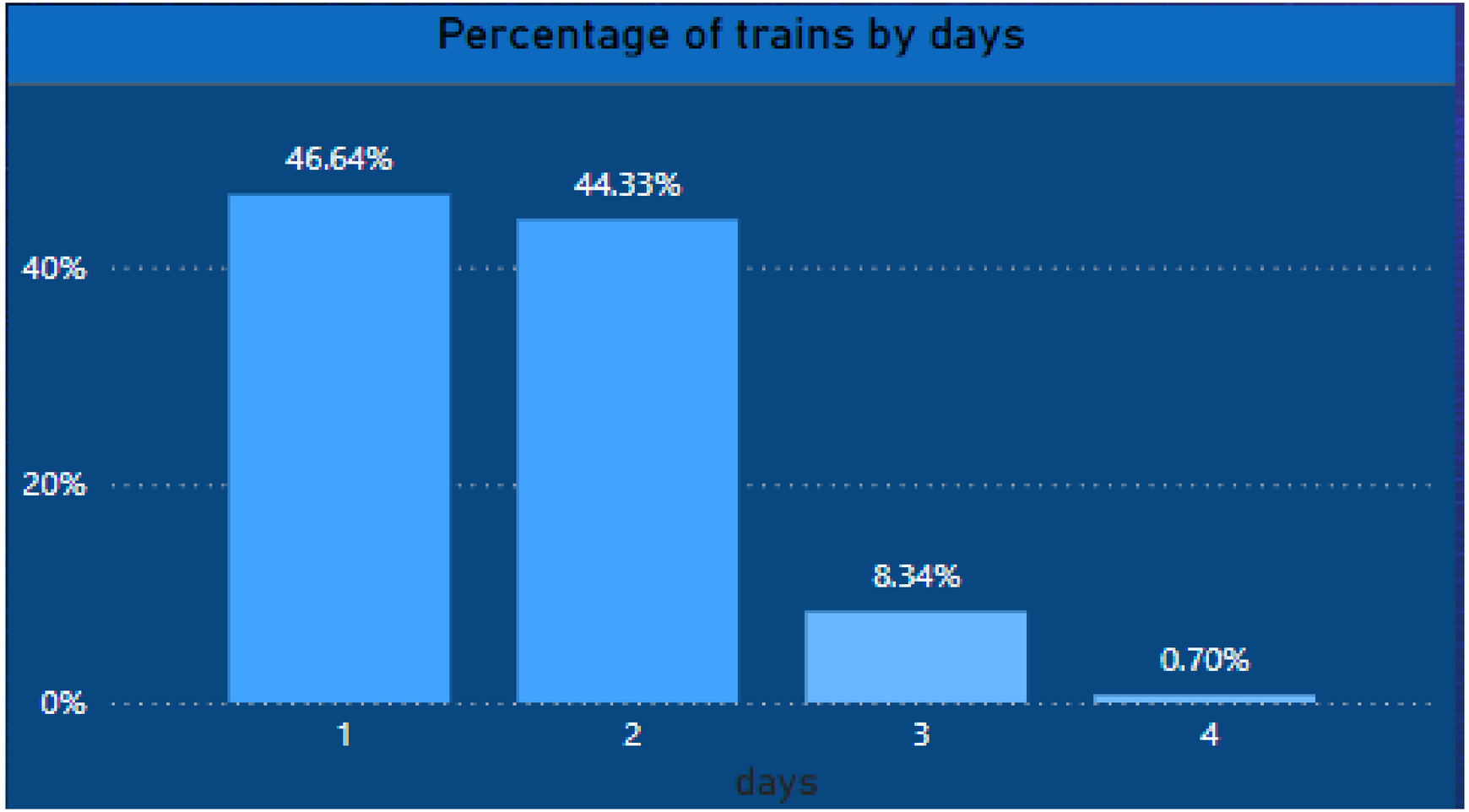
Q11. Which are most expensive trains of India?



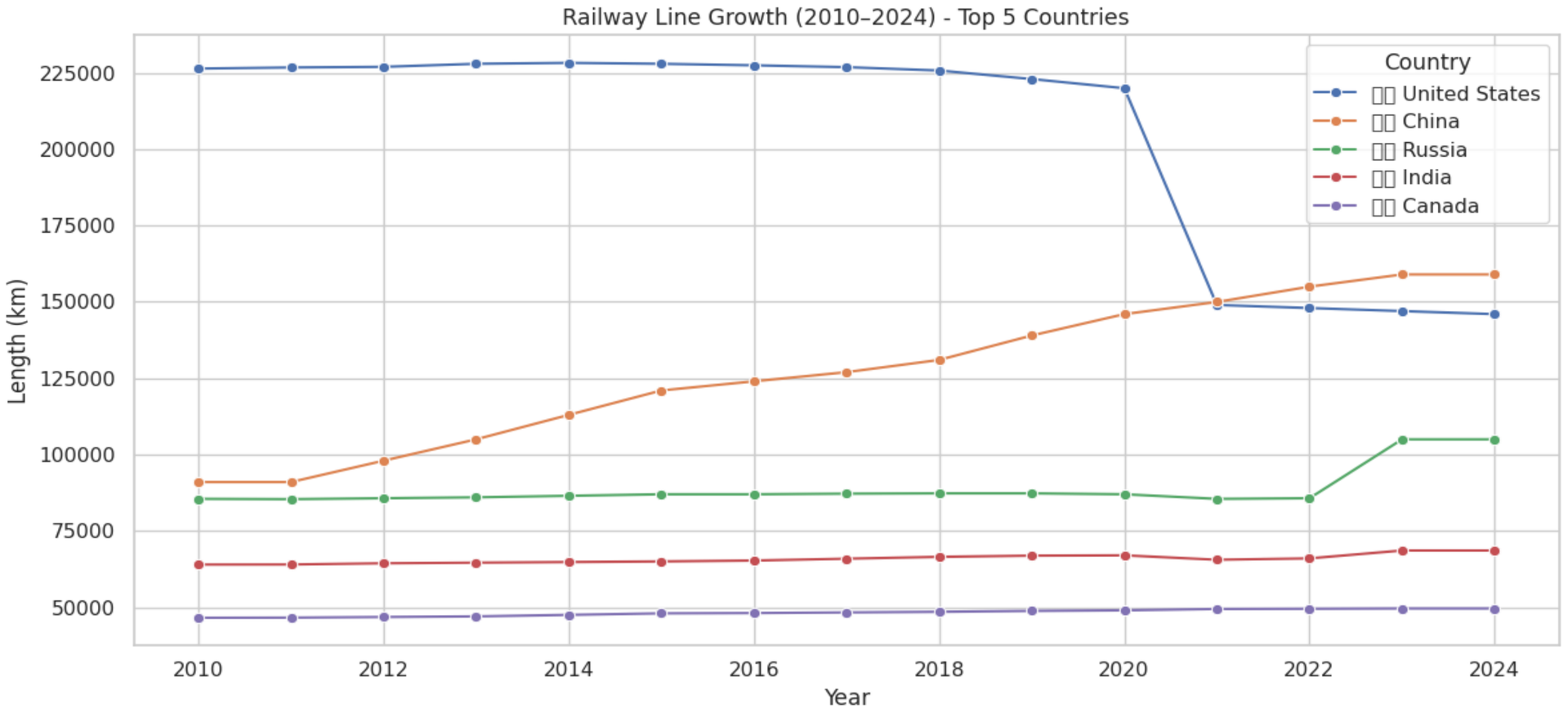
Q12. Which zone have the most trains?



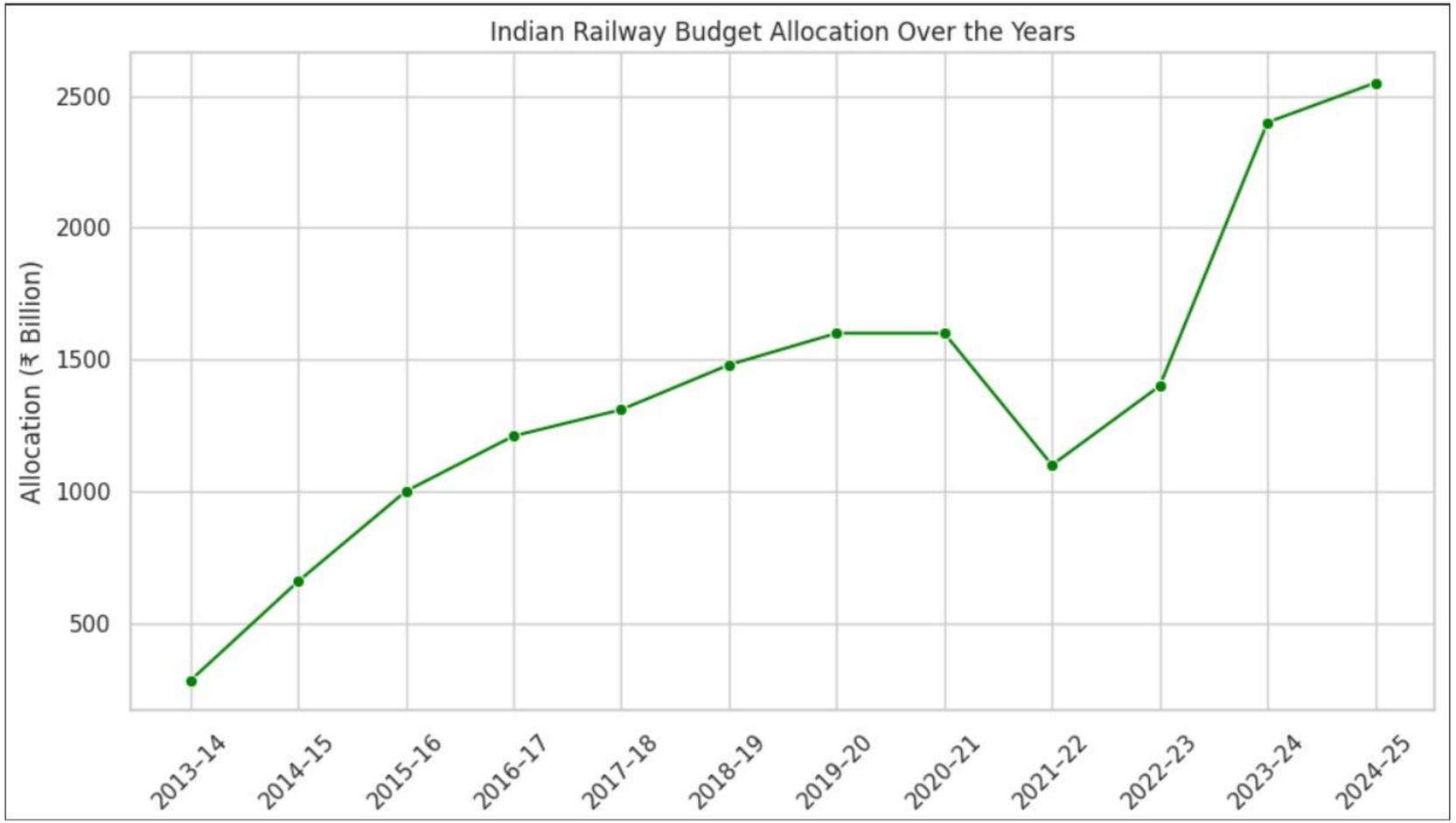
Q13. How much percent of trains run fixed days of week?



Q14. Which are top 10 countries in terms of railway lines?

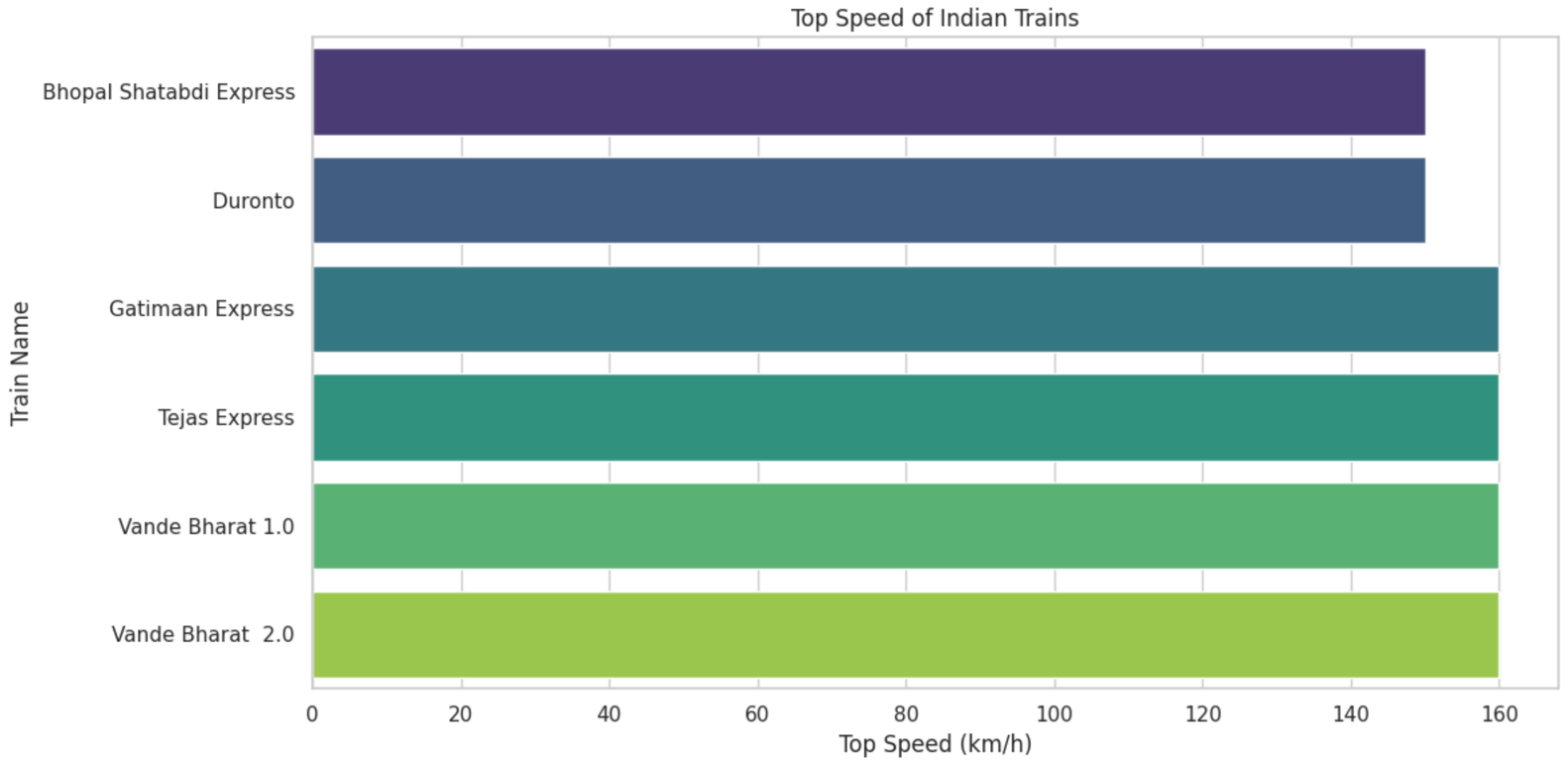


Q15.What is stats of budget allocated to Indian Railways?

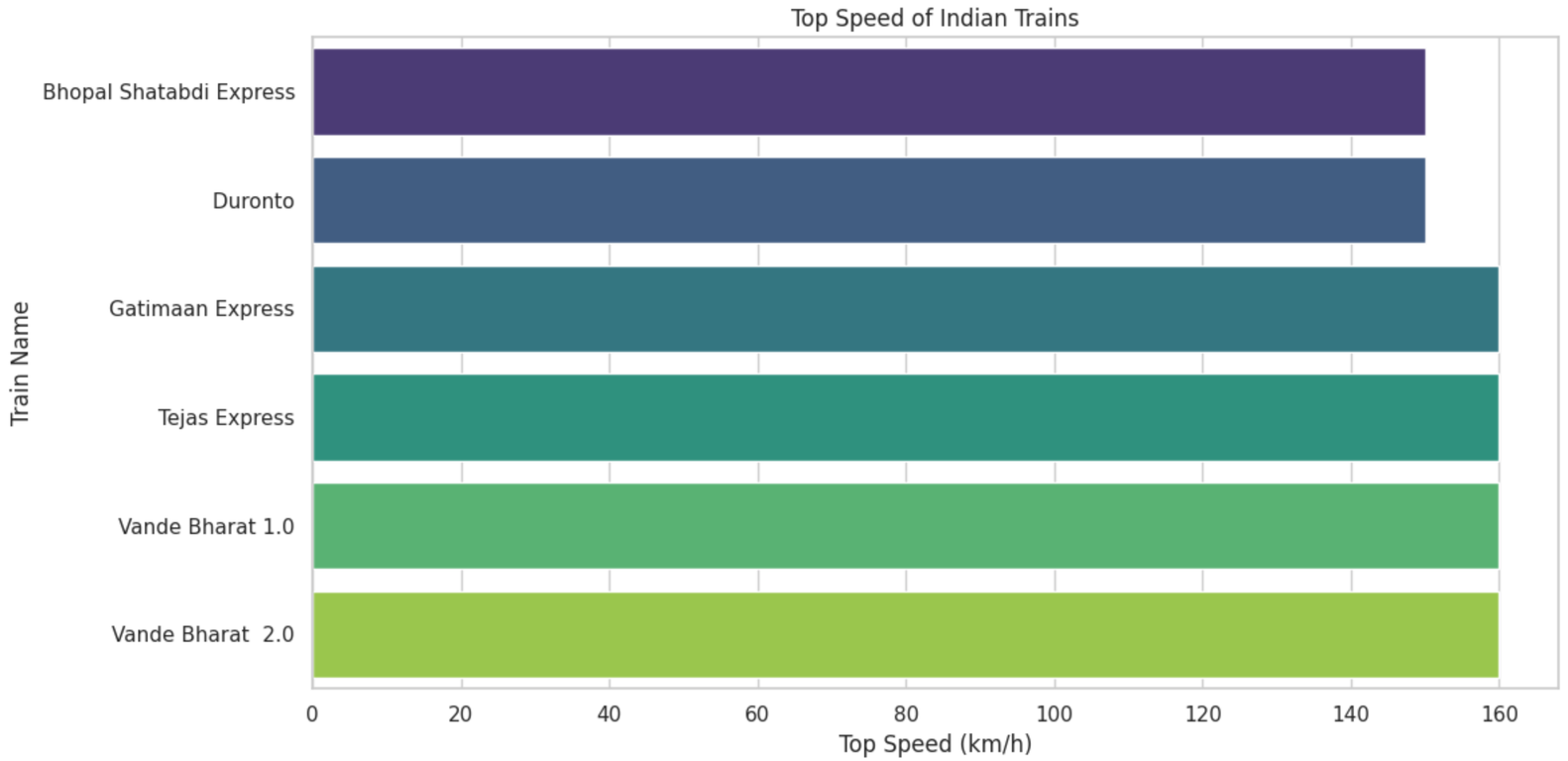


Q16.How many employees work in Indian Railways?

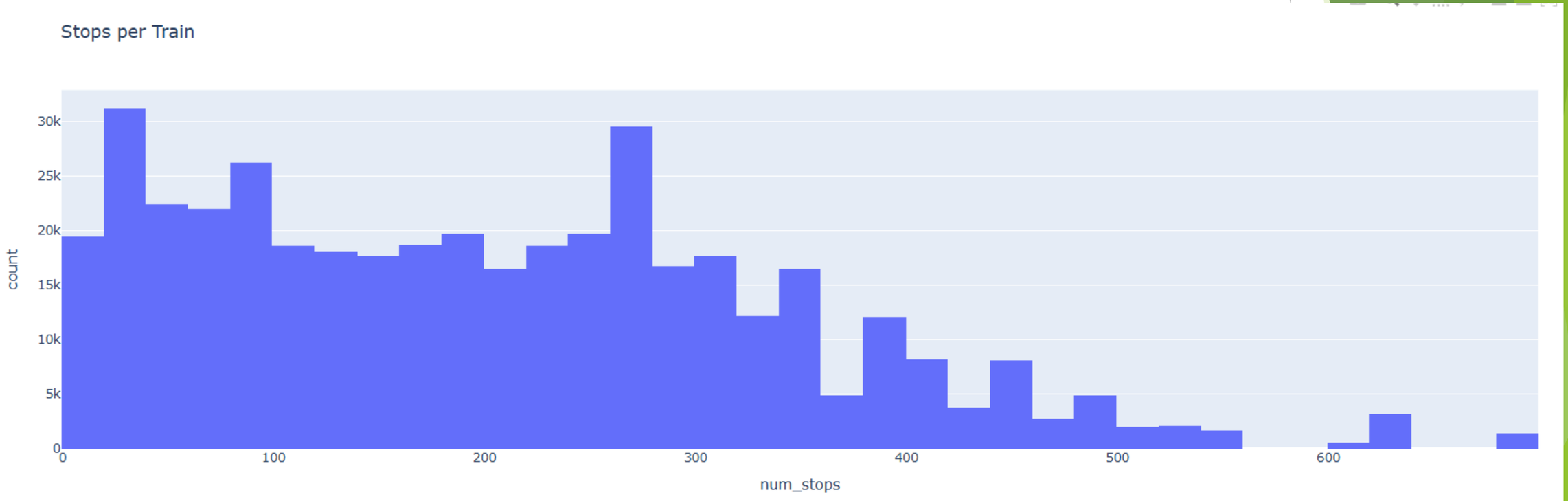
Q17.Fastest Trains of India?



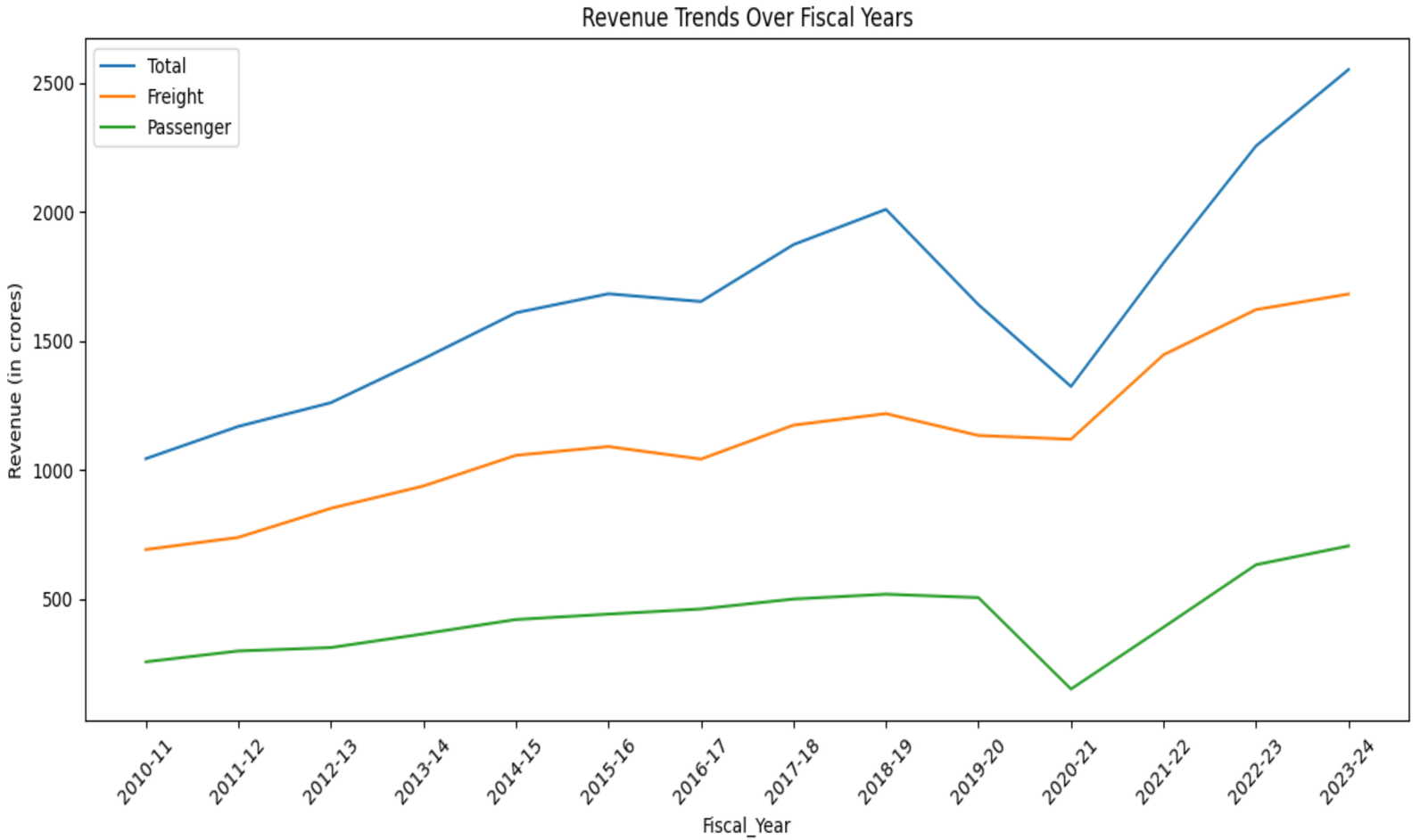
Q18.Fastest Trains of India?



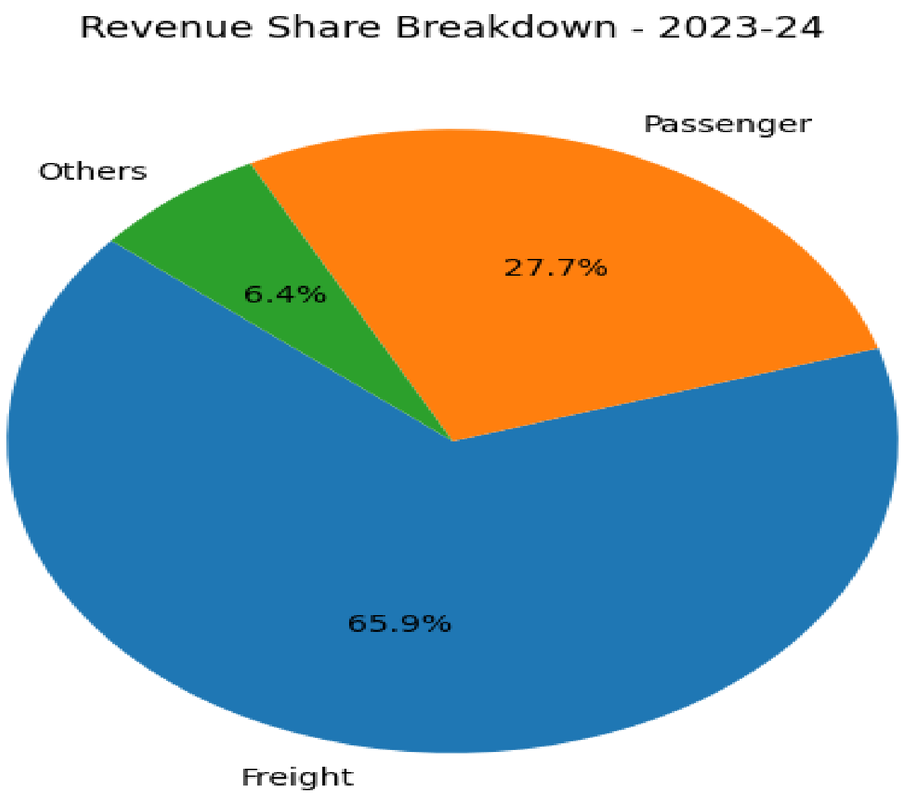
Q19.How many Stops are there Per Train ?



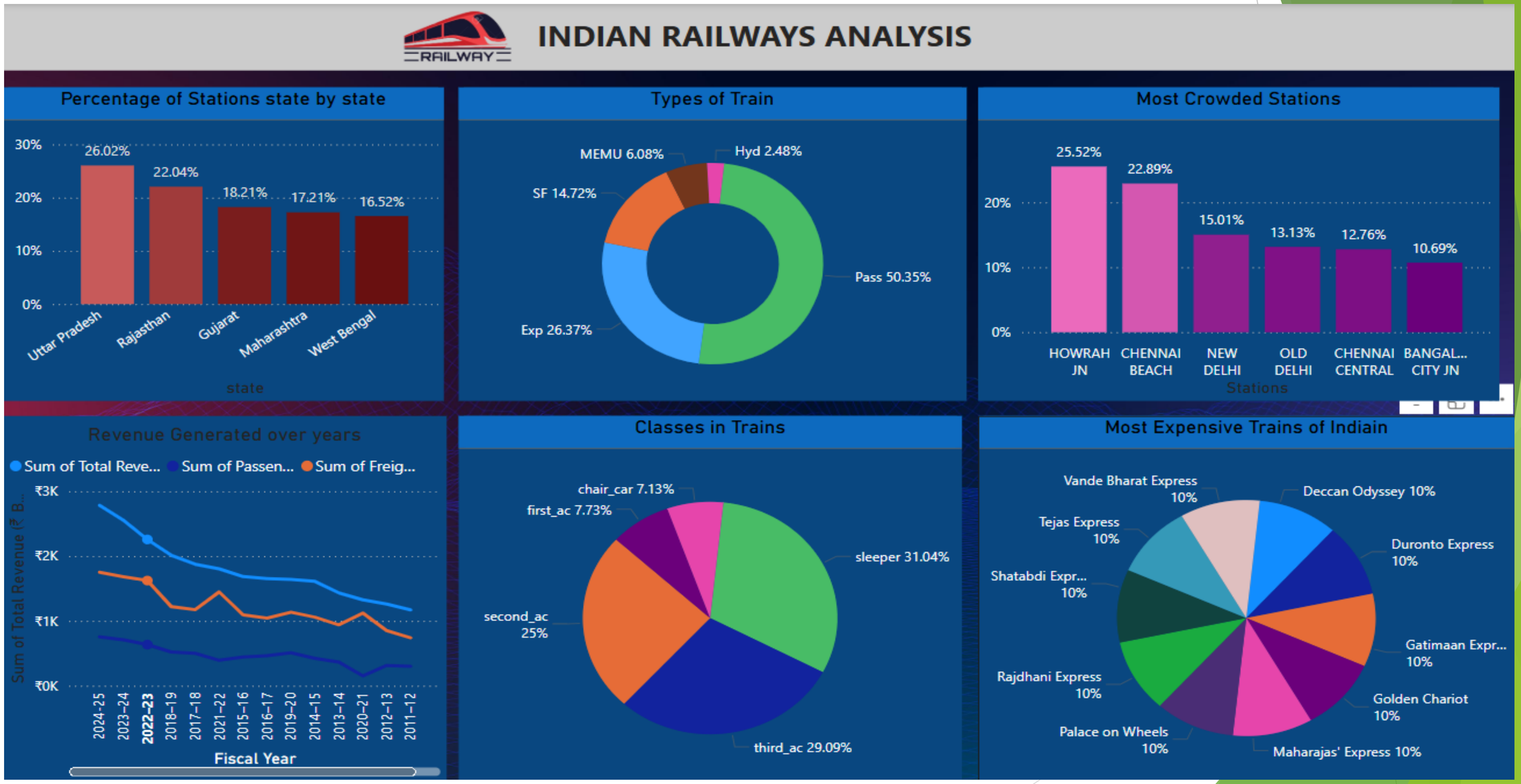
Q20.What are Revenue Trends?



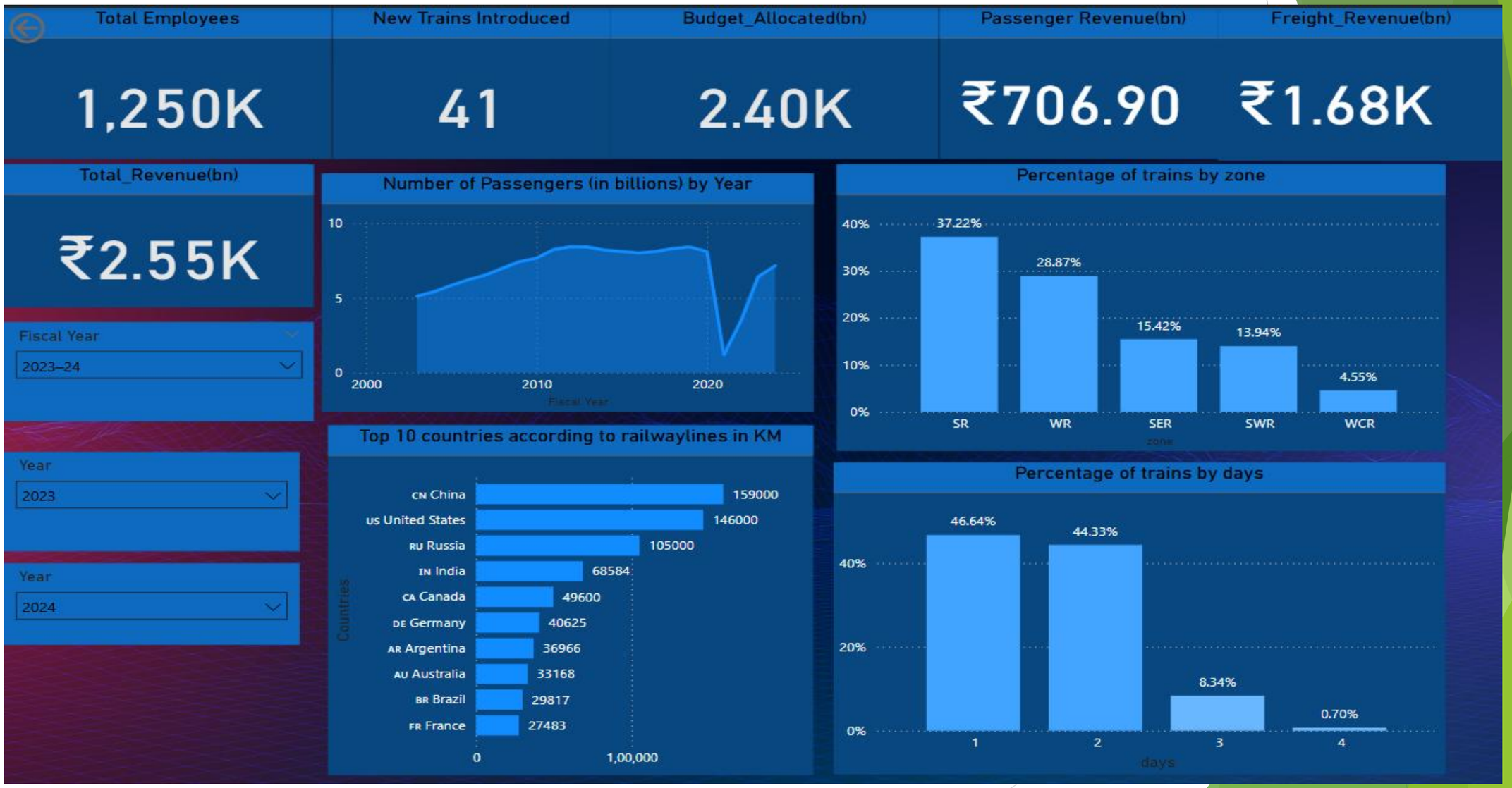
Q21.What is Revenue Share breakdown if2023-24 ?



Power BI Dashboard



Power BI Dashboard



References

1. Ministry of Railways (India)
🔗 https://indianrailways.gov.in

2. Kaggle – Indian Railways Datasets
🔗 https://www.kaggle.com/datasets/sripaadsrinivasan/indian-railways-dataset

3. Open Government Data (OGD) Platform India
🔗 https://data.gov.in

4. Ministry of Statistics and Programme Implementation (MoSPI)
🔗 http://mospi.nic.in

Thank You