K.Sri Manikanta Sadhwik

Tourism Data Analysis Report

1. Dataset Description

1.1 Source:

The dataset used for this project is Tourism_data.csv containing 8,489 records and 5 columns. It captures monthly visitor counts across districts, months, years, and seasons.

1.2 Columns:

- District
- Month
- Visitors
- Year
- Season

1.3 Data Quality:

- The dataset is clean and contains no missing values.
- The 'Visitors' column is numeric and represents monthly visitor counts.
- The structure supports trend analysis across time, regions, and seasons.

2. Operations Performed

2.1 Data Loading and Inspection

- Loaded the dataset into a pandas DataFrame and verified schema.
- Checked for missing values and confirmed numeric conversion for 'Visitors'.
- Computed summary statistics for visitor counts.

2.2 Aggregations and Visualizations

- Grouped data by District and Season to identify top-performing regions.
- Generated charts (line and bar) to visualize trends.
- Applied aggregations and sorting to extract top districts and months.

3. Key Insights

3.1 Overall Visitor Volume

- Total visitors recorded: 715,410,282
- Average visitors per observation: 84,275.0
- Min: 0, Max: 9,761,776

3.2 Top Districts

- Mulugu (77,297,381)
- Hyderabad, Ranga Reddy, Medchal & Vikarabad (61,325,331)
- Yadadri Bhongir (49,289,828)
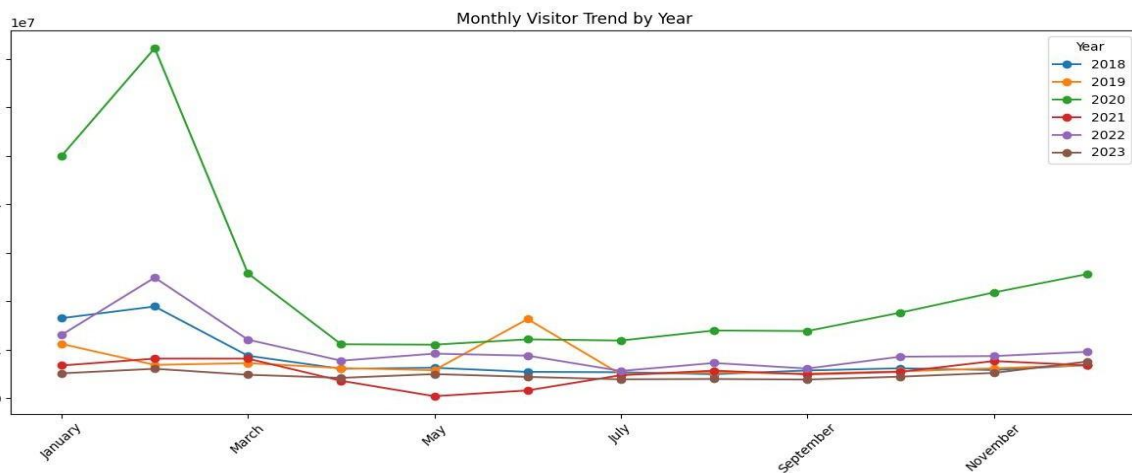- Rajanna Sircilla (44,543,745)
- Hyderabad (43,239,548)

Visitors Heatmap (Top 20 districts by month)
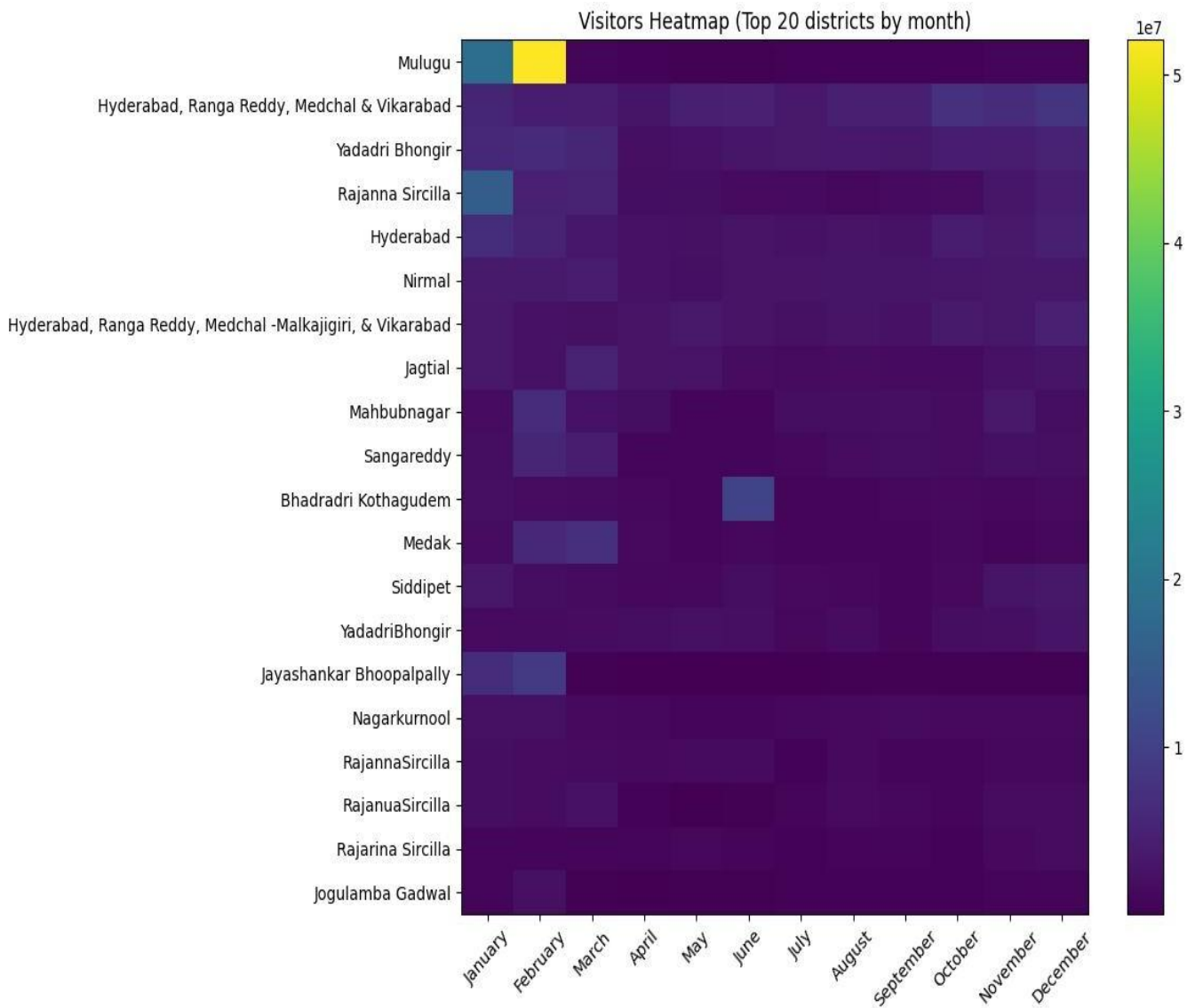
3.3 Seasonal Patterns

- Summer: 196,004,551
- Autumn: 174,939,935
- Winter: 172,453,632
- Monsoon: 172,012,162

Visitor Distribution by Season
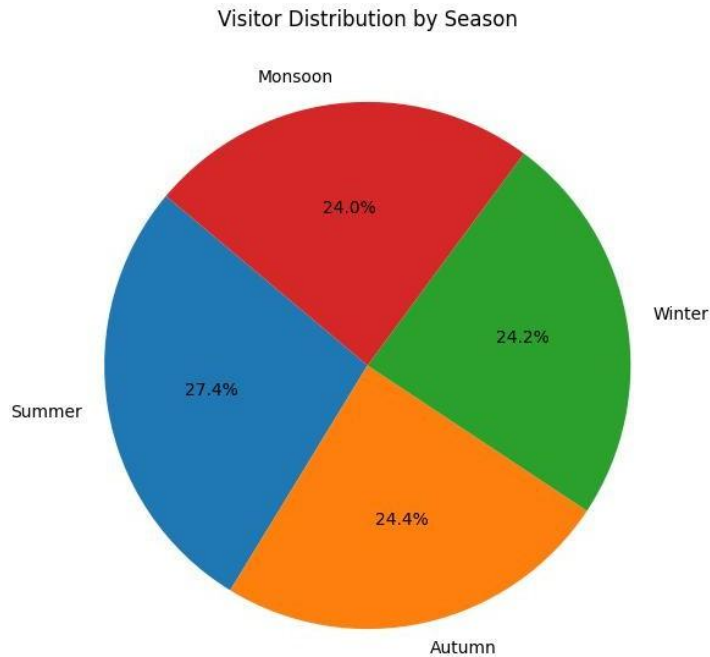
3.4 Monthly Trends

- Top 5 months (by avg visitors): February, January, March, December, November

3.5 Year Coverage

- Dataset spans years 2018 - Winter, supporting seasonal and year-over-year comparisons.

4. Recommendations

4.1 District-Level Focus

- Target marketing for top-performing districts.
- Invest in infrastructure in high-traffic areas.

4.2 Seasonal Readiness

- Allocate staff and resources ahead of peak seasons.

4.3 Monthly Promotions

- Promote during low-average months to balance demand.

4.4 Data Monitoring

- Establish dashboards for real-time district-level monitoring.

4.5 Predictive Modelling

- Use time series or regression models for visitor forecasting.

5. Future Analytics Opportunities

- Develop time-series forecasting models.
- Cluster districts by visitation patterns.
- Combine external data (weather, events, transport) to explain anomalies.

6. Conclusion

- The analysis highlights clear links between season, district, and visitor volumes. Clean, structured data enables reliable aggregation and visualization. Next steps include dashboard creation and predictive modeling to drive data-informed tourism planning.