NAME : S JENESHA MALAR

ROLL NO : 22ADR043

DEPARTMENT : ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Forecasting Public Transport Usage Using Prophet

Key Insights

1. Seasonal Trends in Public Transport Usage

- **Insight**: Passenger numbers vary significantly by season, with higher usage during school terms and lower numbers during holidays (e.g., December and January). For example, "School" service drops to zero during holiday periods, while other services like "Local Route" and "Rapid Route" also show declines.
- **Evidence**: On 28/12/2021, "School" journeys were 0, while on 01/03/2023, they surged to 5,544.

2. Impact of COVID-19 on Transport Demand

- **Insight**: The dataset shows a sharp decline in passenger numbers during 2020 and early 2021, coinciding with pandemic restrictions. Services like "Light Rail" and "Peak Service" were heavily affected.
- **Evidence**: On 20/04/2020, "Local Route" dropped to 379 passengers, and "Peak Service" was 0, compared to pre-pandemic averages of 10,000+ for "Local Route."

3. Peak vs. Off-Peak Service Disparity

- **Insight**: "Peak Service" usage is highly volatile, often dropping to zero on weekends or non-school days, while "Rapid Route" and "Local Route" maintain more consistent numbers.
- **Evidence**: On weekends like 11/09/2021, "Peak Service" was 0, whereas "Local Route" still recorded 820 journeys.

4. Growth in Light Rail Popularity

- **Insight**: "Light Rail" usage has shown a recovery and growth trend post-2021, suggesting increased adoption or expanded service coverage.
- **Evidence**: From 2,352 passengers on 28/12/2021 to 12,226 on 07/03/2024, indicating a 5x increase.

5. Anomalies and Special Events

- **Insight**: Certain dates (e.g., 29/09/2024) show near-zero values for all services, likely due to public holidays or disruptions. Conversely, spikes (e.g., 24/07/2019) may reflect events or promotions.
- **Evidence**: On 29/09/2024, "Local Route" dropped to 1, while on 24/07/2019, it peaked at 19,380.

School Service Dependency

- **Insight**: "School" service is highly dependent on term dates, with predictable spikes during school days and drops to zero during breaks.
- **Evidence**: Consistent zero values in December/January and spikes in March (e.g., 5,544 on 01/03/2023).

Introduction

Forecasting time series data in public transport usage is essential for operational planning and resource allocation. In this project, the **Prophet** algorithm, developed by Facebook, was chosen due to its flexibility in modeling complex seasonal patterns, handling missing data, and producing interpretable forecasts with confidence intervals. Prophet is well-suited for datasets exhibiting strong seasonal effects, such as daily passenger counts influenced by weekdays, holidays, and school terms.

Key Model Parameters and Their Roles

1. daily_seasonality (bool)

- Controls whether daily seasonality is included in the model.
- Useful when data exhibits patterns within the day (e.g., rush hours).
- Set to `True` in this project because public transport demand varies daily.

2. weekly_seasonality and yearly_seasonality

- By default, weekly and yearly seasonal patterns are automatically included.
- Capture effects such as weekend vs. weekday ridership and annual school terms or holidays.

3. changepoint_prior_scale (float)

- Controls the flexibility of the trend's changepoints (where growth rate changes).
- Higher values allow more rapid trend changes (risking overfitting).
- Lower values produce smoother trends.
- Default is 0.05; tuning this affects model sensitivity to sudden shifts.

4. seasonality_prior_scale (float)

- Controls the flexibility of seasonality.
- Higher values allow seasonality to vary more; lower values smooth it out.
- Important to balance fitting seasonal fluctuations without overfitting noise.

5. holidays (DataFrame)

- User-defined dates for holidays or special events.
- Can significantly improve accuracy for irregular demand drops/spikes (e.g., public holidays).

6. growth (str)

- Specifies trend type: `'linear'` or `'logistic'`.
- Logistic growth is suitable when there is a natural saturation point (e.g., max capacity).
- Linear growth assumes unlimited growth or decline.

Advantages of Using Prophet

- Handles missing data and outliers gracefully.
- Incorporates multiple seasonalities and holiday effects explicitly.
- Provides uncertainty intervals, useful for planning under uncertainty.
- Easy to use and tune with interpretable parameters.
- Suitable for daily ridership data influenced by seasonality and external events.