**Bus Ticketing Demand Optimization and Forecasting**

**Project Title:** Bus Ticketing Demand Optimization and Forecasting

**Author:** KARI MAHESH BABU

**Date**: February 2025

**Version**: 1.0

**2. Abstract**

This document presents the project focused on optimizing bus ticketing demand through forecasting models. The primary objective is to improve resource allocation, reduce operational costs, and enhance customer satisfaction by predicting ticket revenue for bus services. The Project leverages time-series forecasting models, including ARIMA, SARIMA, and LSTM, to predict ticket sales and revenue based on historical data. The results aim to assist in better scheduling, reduce overbooking or underutilization, and minimize customer cancellations.

**3. Table of Contents**

1. Title Page
2. Abstract
3. Table of Contents
4. Introduction
5. Literature Review / Background
6. Methodology
   * Data Collection
   * Tools and Technologies
   * Data Preprocessing
   * Model Building
   * Evaluation Metrics
7. Results and Discussion
8. Conclusion

**4. Introduction**

**Problem Statement:**

The bus service struggles with inefficient resource allocation, overbooked or underutilized trips, and customer dissatisfaction due to misaligned schedules and unpredictable demand patterns. These issues lead to revenue loss, increased operational costs, and poor customer experiences, such as frequent cancellations or routes operating with minimal occupancy

**Objectives**:

* Maximize prediction accuracy for feature ticket sales and passenger demand
* Minimize operational costs by optimizing trip schedules and resource allocation

**Constraints:**

* Ensure full compliance with applicable data privacy regulations.
* Maintain the interpretability and usability of the predictive model to facilitate actionable insights and informed business decision-making.

**Success Criteria:**

**Business Success Criteria:**

* Achieve at least a 15% improvement in customer satisfaction scores based on feedback surveys.

**Machine Learing Success Criteria:**

* Achieve a Mean absolute Percentage Error ( MAPE) of less than 10% in demand forecasting.

**Economic Success Criteria:**

* Reduce operational cost ( fuel, staffing) by at least 15% through optimized resource allocation

**Scope:**

This Project covers:

* Data preprocessing (cleaning, handling missing values, encoding, etc.)
* Time-series analysis using ARIMA, SARIMA, and LSTM models.
* Evaluation of models based on performance metrics (RMSE, MAE).
* Providing actionable insights for bus service optimization.

**Methodology:**

**Data set details:**

* The Data set contains 14278 rows and 13 columns

**Data Dictionary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Columns\_Name** | **Data**  **Type** | **Scales of Measurement** | **Description** |
| Date | Date/Time | Interval | Represents the specific date on which the data for a particular bus route is recorded. |
| Bus Route No. | String | Nominal | Denotes the specific bus route identifier. |
| From | String | Nominal | Starting location of the bus route. |
| To | String | Nominal | Destination location of the bus route. |
| Trips Per Day | Integer | Ratio | Number of trips made by the bus in a day. |
| Way | String | Nominal | Indicates whether the bus route is a single trip or includes a return trip. |
| Bus Stops Covered | Integer | Ratio | Total number of bus stops covered on the route. |
| Frequency (mins) | Integer | Ratio | Time interval between consecutive buses on this route. |
| Distance Travelled (km) | Float/Decimal | Ratio | Total distance travelled by the bus during one trip. |
| Travel Time (mins) | Integer | Ratio | Total time taken for the bus to complete one trip |
| Main Station | String | Nominal | The main bus station associated with the route. |
| Tickets Sold | Integer | Ratio | Total number of tickets sold on a single route. |
| Revenue Generated (INR) | Float/Decimal | Ratio | Total revenue generated from ticket sales on the bus route of single trip |

**Tools and Technologies:**

* Programming language-----Python
* Database -------------------- MY SQL
* Visualization tool------------Power BI
* Libraries ------------- Pandas, NumPy, scikit-learn, Statsmodels, TensorFlow
* Models--------- ARIMA, SARIMA, LSTM
* Evaluation Metrics--- RMSE, MAE, MAPE

**3. Data Preprocessing:**

**Steps Performed:**

* **Handling Missing Values:**
  + Checked for null values and imputed missing data appropriately
* **Duplicate Removal:**
  + No duplicate values found in Dataset
* **Data Transformation:**
  + Converted date column to datetime format
  + Type -casting columns to appropriate data types
* **Outlier Detection and Treatment:**
  + Applied Winsorization to handle extreme outliers
* **Scaling:**
  + Used MinMaxScaler for feature Scaling

**4. Forecasting Models:**

**Model Building:**

* **ARIMA (**AutoregressiveIntegrated Moving Average):
  + Applied for linear forecasting based on historical data
* **SARIMA (Seasonal ARIMA):**
  + Used for seasonal data pattern
* **LSTM (Long Short-Term Memory):**
  + Captures complex nonlinear temporal relationships

**Evaluation Metrics:**

**RMSE (Root Mean squared Error):**

* Measures square root of the average squared differences between observed and predicted values.

**MAE (Mean Absolute Error):**

* Represent the average of absolute differences between predicted and observed values

**MAPE (Mean Absolute Percentage Error):**

* Calculate the percentage of deviation between predicted and actual value

**Deployment:**

**Streamlit Application:**

File Name: Bus.py

**Results and performance Analysis:**

**ARIMA:** Suitable for short-term linear forecasting but struggled with seasonal patterns

**SARIMA:** Performed well on capturing seasonal trends

**LSTM**: Provided the most accurate long-term forecasts with lowest error metrics

**Visualization:**

* Model Error Comparison (Bar Chart)
* Predicted vs Actual Revenue (line Plot)

**Conclusion:**

The project successfully demonstrated the use of ARIMA, SARIMA, LSTM models to predict bus revenue, Among the three, LSTM provided most accurate forecasts. This information can help bus operators optimize their schedules and resource allocation, leading to better customer satisfaction and operational efficiency `