Oklahoma City University

Masters in Computer Science

PROJECT

ON

Sales Prediction in the Tourism Industry

Under the guidance of

Tashfeen, Ahmad

Mathematics & Computer Science

BY:

Akhila Mediboyina

Id: B00115521

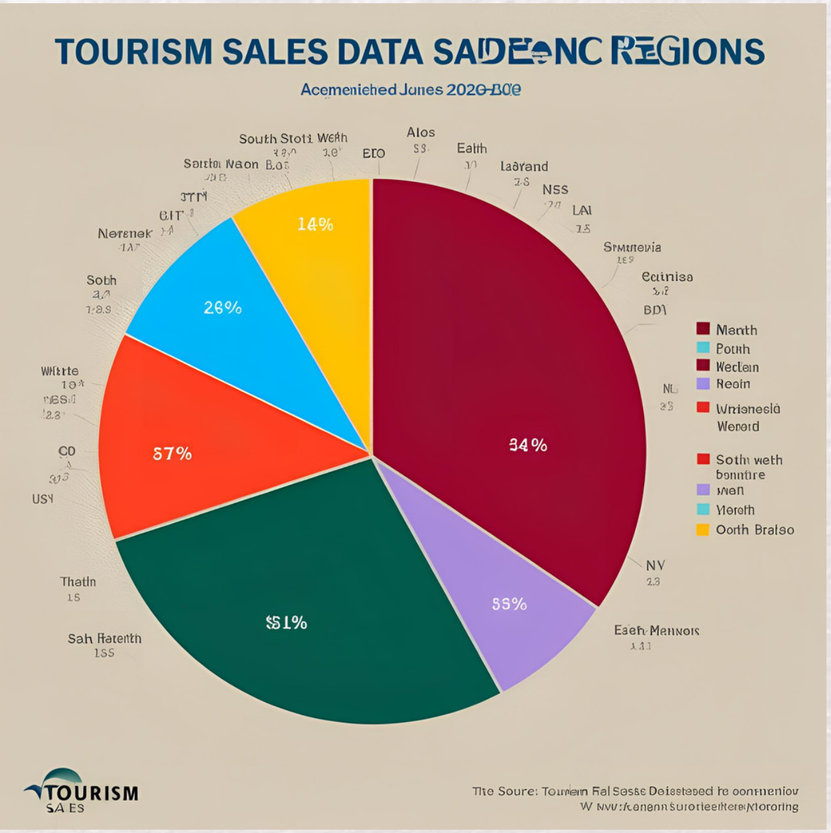
CONTENTS

INTRODUCTION

- **OBJECTIVE:** Our project focuses on sales prediction in the tourism industry, leveraging historical booking and sales data to forecast future trends. By employing advanced machine learning models, we aim to empower stakeholders with actionable insights to optimize revenue, enhance customer experiences, and plan resource allocation effectively.

- **FOCUS AREAS:** Handling missing values, encoding categorical variables, normalizing numerical data, and preparing data for predictive modeling

DATASET OVERVIEW
Tourism Sales Data Across Regions



- **Collector(s)**: Kaggle, World Tourism Organization (UNWTO), Government Tourism Boar
- **Year Range**: 2010–2024
- **Dataset Title**: "Tourism Sales Data Across Regions"
- **Version Number**: v1.2
- **Publisher**: Kaggle, UNWTO, government tourism department
- **Challenges in Raw Data:**
  - Presence of missing values affecting data integrity
  - Categorical data requiring conversion into numerical format
  - Different scales in numerical attributes leading to inconsistencies

DATA PREPROCESSING TECHNIQUES
Preparing Data for Sales Prediction

- **Handling Missing Values**:
  - Identifying and replacing missing values using appropriate strategies (mean, median, mode)
  - Ensuring data completeness to improve prediction accuracy
- **Encoding Categorical Variables**:
  - Transforming non-numeric data into numerical form using techniques like Label Encoding and One-Hot Encoding
- **Normalizing Numerical Data**:
  - Standardizing features to ensure uniform data distribution for better model performance
- **Ensuring Data Quality**:
  - Removing duplicate records and inconsistencies
  - Balancing the dataset to prevent bias in predictions

IMPLEMENTATION HIGHLIGHTS

Key Steps in Data Preprocessing

- **Loading and Cleaning Data**:
  - Read dataset and check for missing values
  - Fill missing values with forward-fill or imputation techniques
- **Transforming Data**:
  - Encode categorical variables (e.g., Region, Tourism Type)
  - Normalize numerical features (e.g., Sales, Revenue)
- **Validating Preprocessed Data**:
  - Visualizing data distributions before and after transformations
  - Ensuring dataset is ready for predictive modeling

CONCLUSION
Summary & Next Steps

- **Key Takeaways**:
  - Preprocessing is critical to improving prediction accuracy
  - Proper handling of missing values and scaling ensures data consistency
  - Transforming categorical data enables seamless model training
- **Next Steps**:
  - Feature engineering to extract additional insights
  - Applying machine learning algorithms for sales prediction
  - Evaluating model performance using suitable metrics


thank you