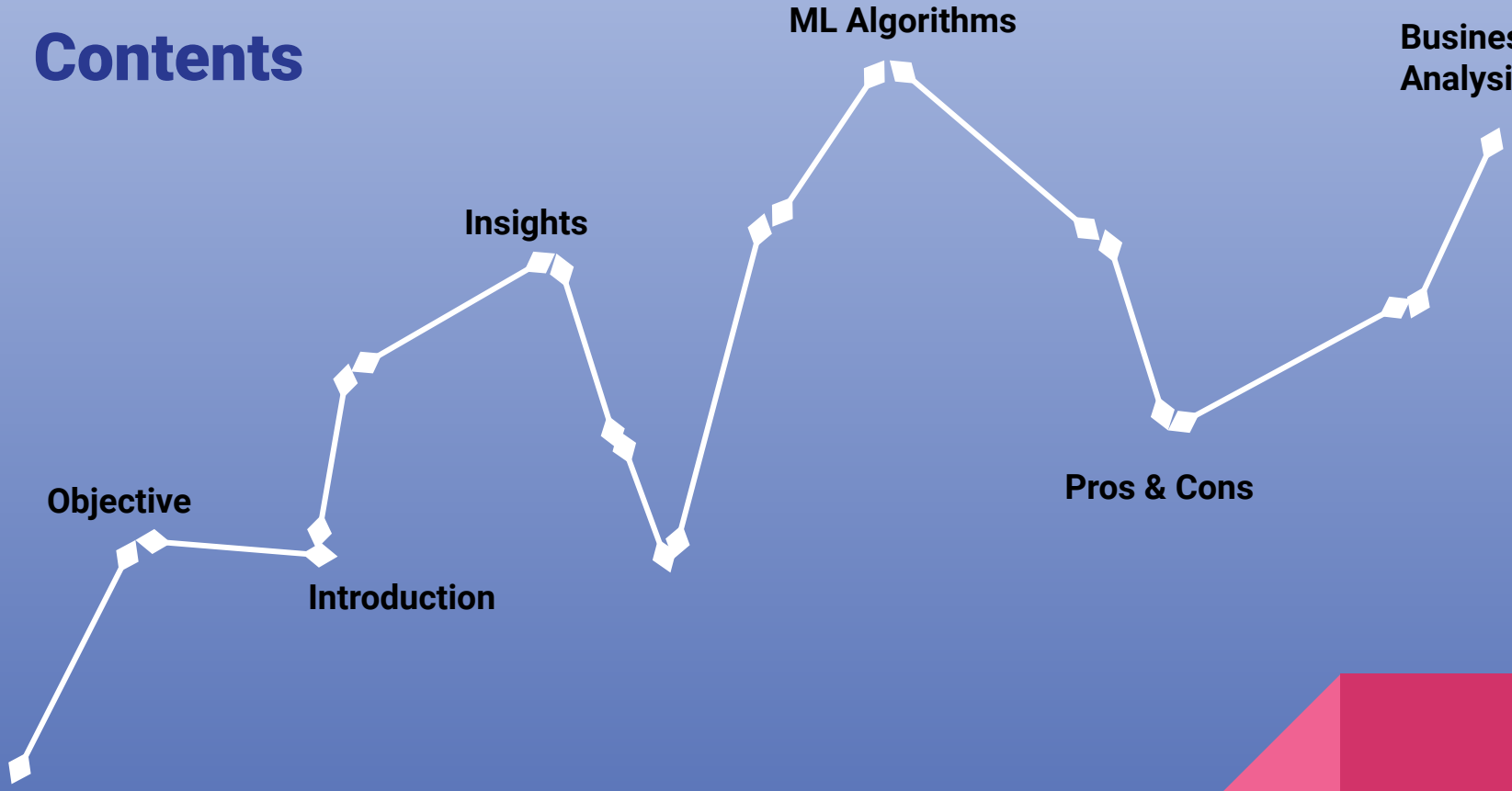
Sales Forecasting

ML Portfolio
Disha Patel
B-17

Contents


ML Algorithms
Business Analysis
Insights
Objective
Introduction
Pros & Cons

Introduction

- Develop SALES FORECASTING MODEL

- The project aims to:
  - Clean and preprocess the dataset to ensure data accuracy and completeness.
  - Explore and visualize the dataset to gain insights into sales patterns and trends.
  - Compare and select the most accurate forecasting model to make future sales predictions with high confidence.
  - Provide clear and concise interpretations of the results and insights gained from the analysis.
  - Overall, project is to showcase the significance of sales forecasting in predicting future sales and aiding the retail store in enhancing its operations and strategy

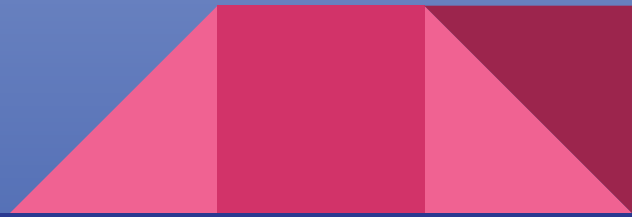
Sales Forecasting

- Sales forecasting is the process of predicting future sales performance based on historical data and other relevant factors.It involves analyzing historical sales data, market trends, economic indicators, and other relevant factors to develop a sales forecast
- Types of sales forecasting methods and techniques may be more suitable for different situations.

1. Qualitative Methods: expert opinions, market research, and are used when historical data is limited or when launching new products or services.
2. Quantitative Methods: methods use numerical data and statistical analysis to predict future sales. Examples of quantitative methods include time series analysis, moving averages, and regression analysis.
3. Hybrid Methods: Combining both Examples include weighted moving averages and collaborative forecasting.

Dataset

- Dataset originally obtained from UCI's online UK retailer
- the date range of the data: 1st Dec 2010 to 9th Dec 2011
- original size of the dataset: 541k rows, 8 columns
- new size of the preprocessed dataset: 235k rows, 12 columns



Insights

- November had the highest sales, possibly due to Thanksgiving and Christmas.
- Sales tend to decrease towards the end of the month, possibly due to paycheck timings.
- More orders are received from Monday to Thursday and decrease afterward.
- Most orders are received at 12:00pm, possibly during lunch hour.
- Ireland has the highest average spend per customer, while the UK has the highest total sales.

Feature Selection and Data Preprocessing

SS of sales df and explain to do data preprocess

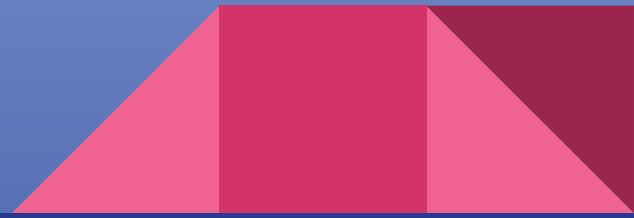
Models

Baseline models

Final Model LSTM

RNN model

Most frequently used for Time-Series forecasting



Results From All Models

Business Interpretation

Pros and Cons

XGBoost:

Pros:

- **Powerful machine learning algorithm**
- **Easy to use**
- **Efficient on large datasets**
- **Provides feature importance scores**

XGBoost:

Cons:

- **Hyperparameter tuning required**
- **Difficult to interpret**
- **Not well-suited for time-series data**

LSTM

Pros:

- **Captures long-term dependencies and trends**
- **Can be used for both univariate and multivariate time-series forecasting**

LSTM

Cons:

- **Requires large amounts of data**
- **Prone to overfitting if not regularized properly**
- **Computationally expensive to train**
- **Difficult to interpret**

Conclusion

- Sales forecasting methods can be debatable and is a complex process that requires domain expertise .
- It can helps businesses to plan, make informed decisions, and manage their operations effectively and enable them to optimize their inventory, production, marketing, and other aspects of their operations and maintain profitability over the long term.
- By analyzing historical sales data and other relevant factors, businesses can develop accurate sales forecasts that allow them to identify and reduce potential risks and capitalize on opportunities.

Next steps

- Gathering Quality data relevant to time-series sales
- Adding Time-series Features such as lag, rolling window
- Incorporating traditional techniques such as ARIMA and creating hybrid model for better forecast
- With more computation power adding stacking and with good understanding of deep learning specific to Time-Series forecast