High Level Design

Driving Business Insights Through Predictive Modeling

Presented By	Ayan Biswas
Date	23.01.2025

Project Overview

Objective:

- To analyze the Sample Superstore dataset and identify key business insights.
- Build predictive models to enhance sales performance and minimize losses.

Dataset Description:

- Source: [Insert Source]
- Contains sales data, including categories like Region, Segment, Profit, and Discounts.

Key Deliverables:

- Exploratory Data Analysis (EDA).
- Feature Engineering and Preprocessing.
- Predictive Modeling.
- Business Recommendations.

Exploratory Data Analysis (EDA)

Insights:

1. Sales and profit distribution across regions and categories.
2. Impact of discounts on profitability.
3. Identification of high-profit and high-loss segments.

Visuals:

- Sales vs. Profit by Region (Bar Chart).
- Discount vs. Profit (Scatter Plot).
- Segment-wise Profitability (Pie Chart).

Data Preprocessing

Steps Taken:

- Handled missing values using [method].
- Converted categorical data to numerical using one-hot encoding.
- Removed outliers using [technique].
- Standardized numerical features.

Tools:

- Python (Pandas, NumPy).
- Scikit-learn.
- Flask
- Seaborn
- Matplotlib
- Git

Predictive Modeling

Problem Statement:

- Predict sales performance and identify factors driving profitability.

- **Models Used:**

1. Linear Regression (Baseline).
2. Ridge
3. Lesso
4. Random Forest Regressor.
5. Decition tree regressor.

- **Evaluation Metrics:**

- RMSE (Root Mean Squared Error).
- R^2 Score.
- Accuracy score

- **Results:**

- Best Model: [Random Forest Regressor].
- Accuracy: [63].

Key Findings

1. Discounts above a threshold negatively impact profitability.
2. The "Technology" category yields the highest profit margins.
3. Central and West regions have potential for growth.

Business Recommendations

Short-term:

- Optimize discounting strategies.
- Focus marketing efforts on high-profit regions.

Long-term:

- Expand the "Technology" category offerings.
- Implement dynamic pricing strategies.

Tools and Technologies

- **Languages:** Python
  Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, XGBoost
  Visualization: Tableau, Power BI

Conclusion

- Enhanced decision-making through data-driven insights.
- Successful implementation of predictive models to improve sales and profitability.
- Scope for further analysis: Customer segmentation and lifetime value prediction.

Thank You

- **Contact:** biswas.ayan1997@gmail.com
- **Questions?**