IBM Data Science Capstone project

Dadmehr Rahbari

Executive Summary

Objective: Develop a comprehensive data science project to analyze and predict outcomes based on a given dataset.

Scope: Conduct data collection, wrangling, exploratory data analysis (EDA), predictive modeling, and visualization.

Outcome: Provide insights, actionable recommendations, and a predictive dashboard for stakeholders.

Introduction

Goal: Understand and solve a business challenge through data-driven insights.

Dataset

Tools: Python, SQL, Plotly, Folium, Pandas, NumPy, Scikit-learn, SQL, Jupyter Notebook.

Approach: Data Cleaning → Exploratory Analysis → Predictive Modeling → Dashboard Creation

Data Collection and Wrangling Methodology

Sources: Web APIs, Databases, CSV Files, etc.

Process:

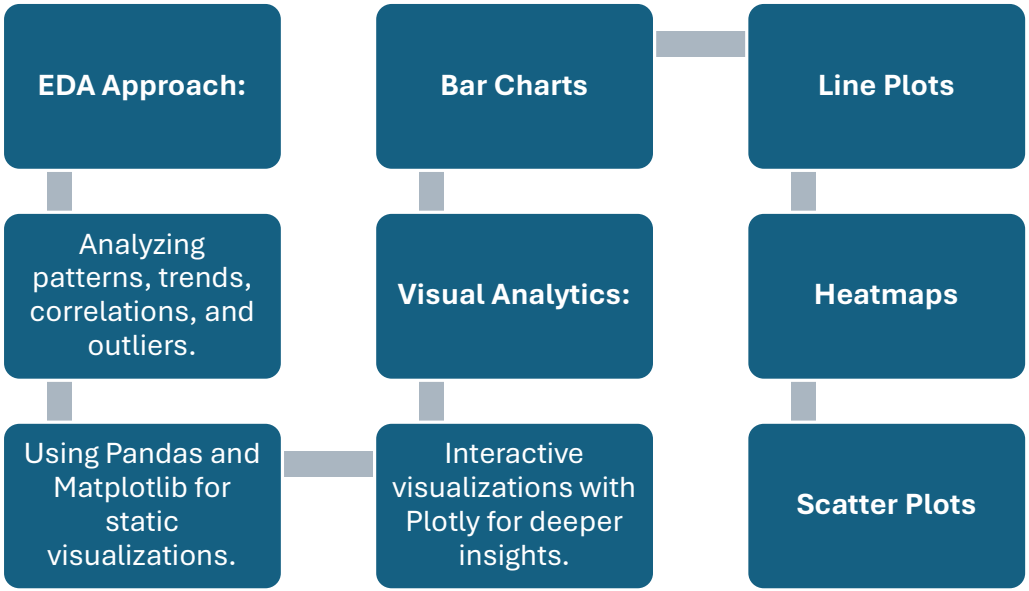
Data Acquisition

Data Cleaning

Data Transformation

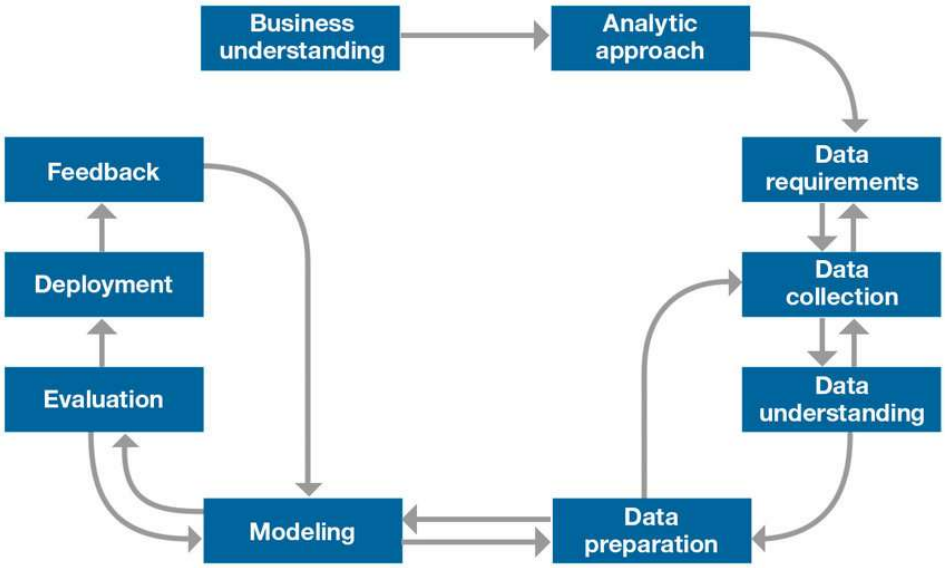
Feature Engineering

EDA and Interactive Visual Analytics Methodology



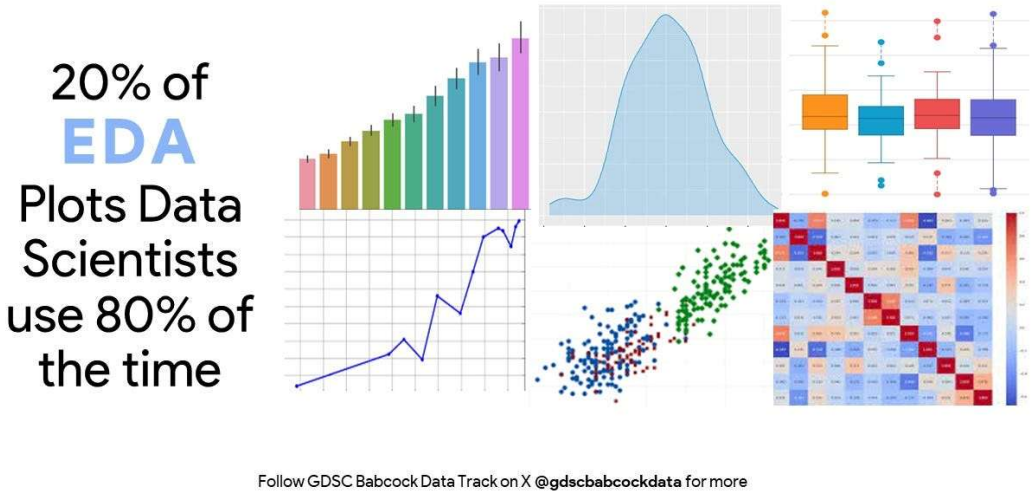
Predictive Analysis Methodology

- **Objective**: Build predictive models for classification tasks.
- **Techniques**: Logistic Regression, Random Forest, and Neural Networks.
- **Evaluation Metrics**: Accuracy, Precision, Recall, F1-score.
- **Tools**: Scikit-learn, TensorFlow, XGBoost.



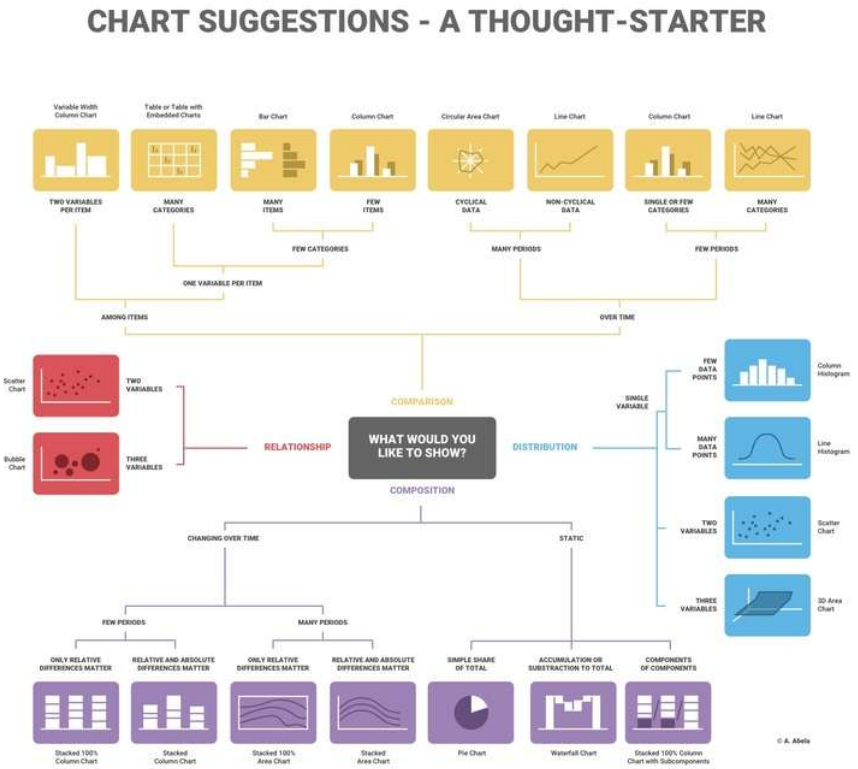
EDA with Visualization Results

- **Distribution of Variables**
- **Correlation Heatmaps**
- **Outlier Detection**
- **Trend Analysis**
- **Comparative Visualizations**
- **Feature Importance Analysis**



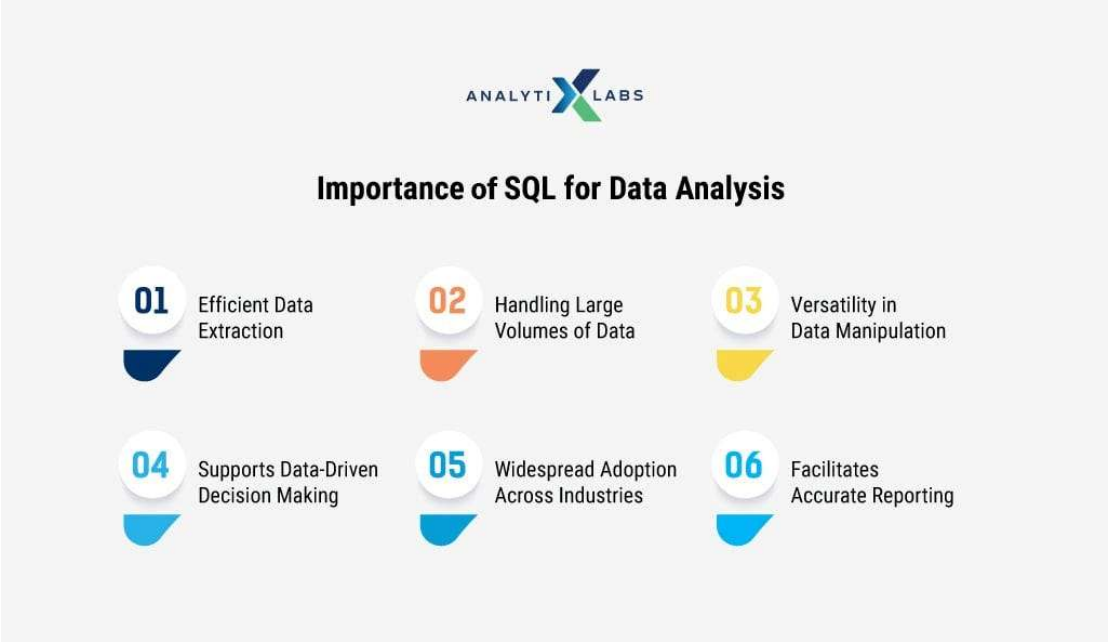
EDA with SQL Results

- Data Query
- Aggregated Metrics
- Joins and Filtering
- Data Summarization
- Data Cleansing
- Performance Analysis
- Complex Queries
- Database Insights
- SQL Optimization
- Results Visualization



Interactive Map with Folium Results

- **Objective**: Visualize geographic data and locations using Folium.
- **Features**:
  - Map Markers
  - Heatmaps
  - Zoom Levels
  - Layered Maps
- **Use Case**: Display geospatial insights related to business operations.



Plotly Dash Dashboard Results

•**Features**:

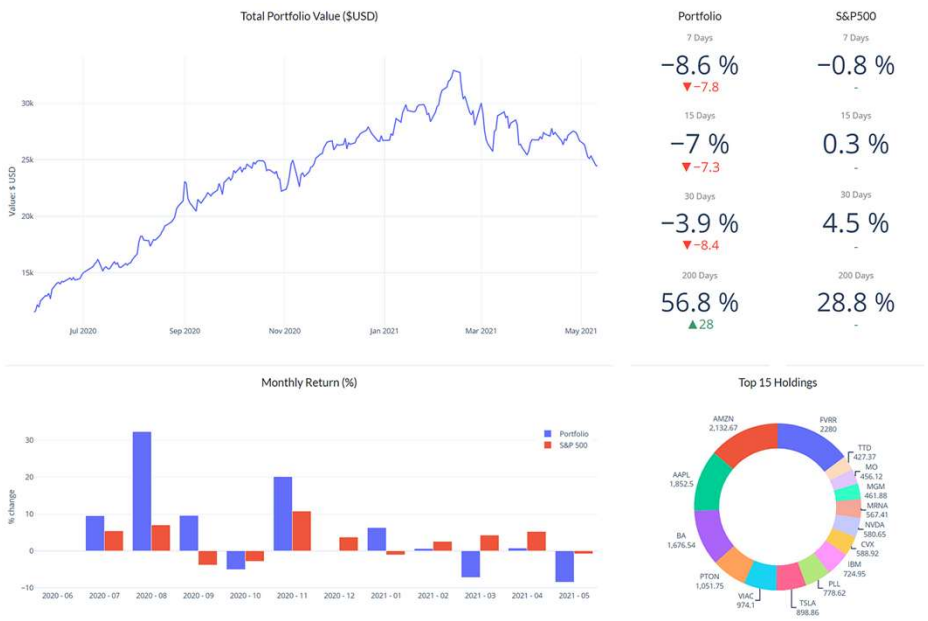
1.Interactive Data Visualizations

2.Real-time Data Updates

3.Customizable Widgets (Filters, Dropdowns)

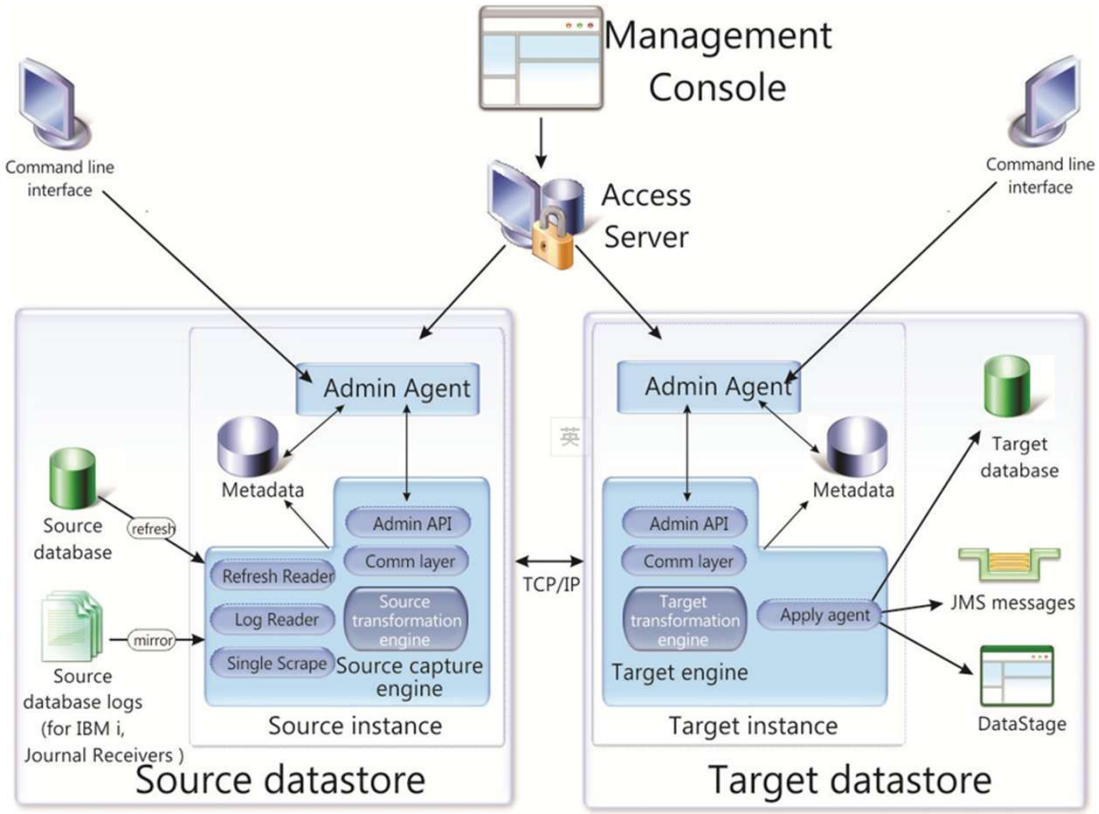
4.Summary Statistics

•**Use Case**: Enable decision-makers to interact with data for trend identification and prediction.



Predictive Analysis (Classification) Results

- Confusion Matrix
- ROC Curve
- Feature Importance
- Prediction Accuracy
- F1-Score and Precision-Recall Trade-off



Conclusion

Summary of key findings and insights.

Recommendations based on predictive and visual analytics.

Next steps for implementation and improvement.