

# Executive Summary

- **Objective:** The objective of this project is to analyze and predict outcomes based on the dataset, leveraging data science tools like Python, machine learning, and visualization techniques.
- **Methods Used:** We applied Exploratory Data Analysis (EDA), machine learning models (e.g., classification), and visualization tools such as Plotly, Matplotlib, and Folium.
- **Main Findings:** Key findings include trends in

# Introduction

- Context: This project analyzes [mention the dataset, e.g., customer data, sales data, etc.] to uncover meaningful patterns and predictions.
- Objective: The goal of the analysis was to predict customer churn, sales trends, or any other specific outcome.
- Dataset: The data was obtained from [mention the source, e.g., Kaggle, a company database, etc.].

# Data Collection and Wrangling

- Data Collection: The dataset was collected from [mention source]. It includes variables such as [list some of the key variables, e.g., age, salary, region].
- Data Wrangling:
  - - Handled missing values using [methods, e.g., mean imputation, dropping missing data].
  - - Removed outliers using [methodology, e.g., Z-score or IQR].
  - - Transformed categorical variables into

# EDA and Interactive Visual Analytics Methodology

- Exploratory Data Analysis (EDA):
  - - Examined the distribution of key features using histograms and box plots.
  - - Investigated correlations between variables using heatmaps.
- Interactive Visual Analytics:
  - - Used Plotly for interactive graphs such as scatter plots, bar charts, and line charts to better understand the trends.
  - - Integrated interactive tools like dropdowns

# Predictive Analysis Methodology

- Machine Learning Models Used:
  - - Logistic Regression, Random Forest, and K-Nearest Neighbors (KNN) to predict outcomes based on the dataset.
- Evaluation Metrics:
  - - Accuracy, precision, recall, and F1 score to ensure reliable predictions.
- Process:
  - - The data was split into training and testing sets using `train_test_split` from `scikit-learn`

# EDA with Visualization Results

- Distribution of Age:
  - - A histogram of the age distribution showed that most individuals are between 25 and 45 years old.
- Correlation Matrix:
  - - A heatmap revealed that feature X and feature Y have a strong positive correlation.
- Visualizations:
  - - Bar charts for categorical features.
  - - Scatter plots showing relationships between

# EDA with SQL Results

- SQL Queries:
- Query 1: Extracted records of customers who purchased more than \$1000 in the last month:
- `SELECT * FROM sales_data WHERE purchase_amount > 1000;`
- Query 2: Aggregated sales by region:
- `SELECT region, SUM(sales) FROM sales_data GROUP BY region;`
- Results:
- SQL helped in filtering and aggregating the

# Interactive Map with Folium

## Results

- Folium Map:
- - Created an interactive map to show customer locations.
- Code Example:
- `import folium`
- `map = folium.Map(location=[latitude, longitude], zoom_start=10)`
- `folium.Marker([latitude, longitude], popup='Customer').add_to(map)`
- Result:



# Plotly Dash Dashboard Results

- Dash Dashboard:
- - Developed an interactive dashboard with Plotly Dash to visualize key metrics such as sales trends, customer demographics, and model performance.
- Features:
- - Includes dropdowns to filter by region, line graphs for sales trends, and bar charts for customer segmentation.

# Predictive Analysis (Classification) Results

- Model Performance:
  - - Logistic Regression: Achieved an accuracy of 85%.
  - - Random Forest: Provided a higher accuracy of 90%.
- Evaluation Metrics:
  - - Show confusion matrix, ROC curve, and classification report for each model.

# Conclusion

- Key Insights:
  - - Our analysis showed that Feature X significantly influences the target variable.
  - - The Random Forest model provided the best classification performance with an accuracy of 90%.
- Actionable Recommendations:
  - - Based on the analysis, we recommend targeting high-value customers in regions with low sales.

# Creativity Beyond the Template

- Enhanced Visuals:
  - - Added custom charts, color-coded heatmaps, and interactive graphs.
- User Experience:
  - - Focused on creating intuitive dashboards that provide meaningful insights at a glance.

# Innovative Insights

- Unexpected Patterns:
  - - We discovered that Feature Z has a much stronger influence on the target variable than initially thought.
- Potential Business Impact:
  - - This finding could be used to optimize marketing strategies targeting specific customer segments.