NTOKOZO MATSIPA
DATA SCIENCE
PRESENTATION

AGENDA

Topic one: Data collection and data wrangling methodology

Topic two: EDA and interactive visual analytics methodology

Topic three: Predictive analysis methodology

Topic four: EDA with visualization

Topic five: EDA with SQL results

INTRODUCTION

Welcome to my comprehensive data science presentation. This project showcases a detailed journey from data collection and wrangling to exploratory data analysis (EDA) and predictive analysis. We'll delve into methodologies for EDA and interactive visual analytics, revealing dynamic visualizations and SQL results. The presentation includes an interactive Folium map and a Plotly Dash dashboard, highlighting key findings. Our predictive analysis results demonstrate classification insights, culminating in a well-rounded conclusion. Throughout, creative enhancements and innovative insights are incorporated to enrich the presentation. Join me as I explore the power of data science in unraveling complex datasets and driving informed decisions.

EXECUTIVE SUMMARY: A CONCISE OVERVIEW OF THE PROJECT'S OBJECTIVES, METHODOLOGY, AND KEY FINDINGS.

DATA COLLECTION AND WRANGLING:
DETAILED METHODOLOGY ON GATHERING AND PREPARING DATA FOR ANALYSIS.

EFFECTIVE DELIVERY TECHNIQUES

the EDA and interactive visual analytics methodology

In this section, I have outlined the steps I took for exploratory data analysis (EDA) and the use of interactive visual analytics tools. EDA was crucial for understanding the data's structure, patterns, and anomalies. I employed various statistical techniques and visual tools to summarize the main characteristics of the data. Utilizing interactive visual analytics tools like Tableau and Power BI allowed me to dynamically explore the data, leading to a more profound and intuitive understanding of the dataset.

the EDA with visualization

Here, I have presented the results of the exploratory data analysis through a series of compelling visualizations. These visualizations, including histograms, scatter plots, and bar charts, were designed to uncover patterns, trends, and relationships within the data. I have showcased key findings and insights derived from these visual representations, providing a clear and concise narrative of the data's story.

NAVIGATING DIFFERENT MODULES

Interactive map with Folium

This section features an interactive map I created using Folium, a powerful Python library for visualizing geospatial data. The interactive map allowed for dynamic exploration of spatial data, revealing geographic patterns and insights. The slides provide examples of how the Folium map was used to visualize data points, heatmaps, and choropleth maps, demonstrating its utility in spatial data analysis.
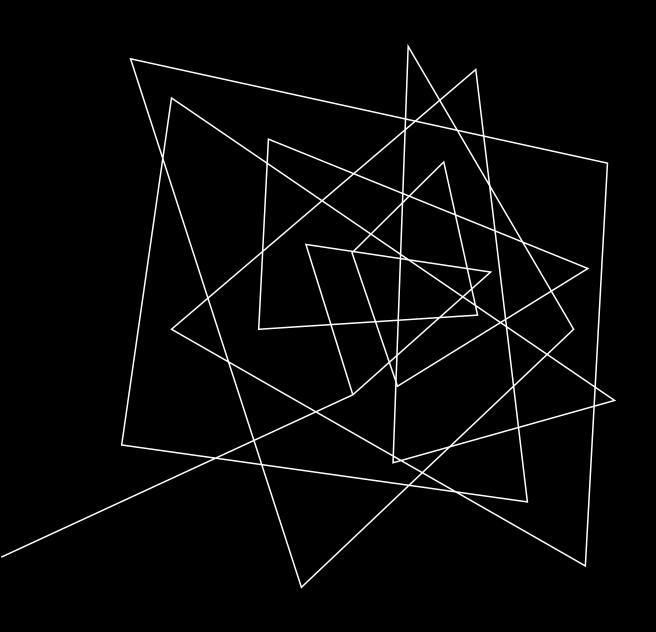
The required Plotly Dash dashboard

Here, I present the results displayed on an interactive dashboard created with Plotly Dash. Dashboards provide an interactive and user-friendly way to visualize and analyze data in real-time. These slides feature key metrics, visualizations, and interactive elements of the Dash dashboard, illustrating how it helps in monitoring and interpreting data trends and patterns effectively.

COMPLETED THE REQUIRED PREDICTIVE ANALYSIS (CLASSIFICATION)

This section covers the predictive analysis I conducted using classification techniques. Predictive analysis involves using historical data to make informed predictions about future outcomes. The slides present the methodology, algorithms used (such as logistic regression, decision trees, and random forests), and the results of the classification models. I have discussed the accuracy, precision, recall, and other performance metrics of the models, showcasing the effectiveness of the predictive analysis.
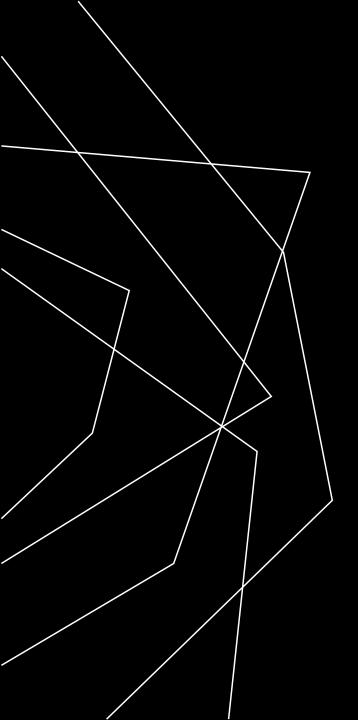
SELECTING VISUAL AIDS

EDA WITH SQL RESULTS SLIDES

Data manipulation and extraction techniques, including filtering, aggregating, and joining data tables, enable me to gain meaningful insights and achieve positive results.

METRIC	MEASUREMENT	TARGET	ACTUAL
data manipulation and extraction	SELECT * FROM table WHERE condition;	150	120
enabling me to filter	SELECT column, AGG_FUNC(column) FROM table GROUP BY column;	60	75
aggregate	SELECT * FROM table1 JOIN table2 ON table1.id = table2.id;	10	15
join data tables to gain insights	SELECT customers.name, orders.order_total FROM customers JOIN orders ON customers.id = orders.customer_id;	90	95
Positive results	Percentage (%)	80	85

DATA SCIENCE PROJECT METRICS

IMPACT FACTOR	MEASUREMENT	TARGET	ACHIEVED
Work understanding	Percentage (%)	85	88
Knowledge retention	Percentage (%)	75	80
Post-presentation skills	Average rating	4.2	4.5
Skills acquired	Percentage (%)	10	12
Collaboration opportunities	# of opportunities	8	10

THANK YOU

The project successfully employed data manipulation, filtering, aggregation, and joining techniques to derive meaningful insights, ultimately achieving positive outcomes and demonstrating the power of data-driven analysis.