Final Capstone Project Presentation

IBM Data Science Professional Certificate

Executive Summary

- This capstone project demonstrates the application of data science methodology to solve real-world problems. It includes data collection, wrangling, exploratory analysis, modeling, and visualization using Python tools.

Introduction

- In this project, various data science skills were applied across several domains, including housing price prediction, stock analysis, and rainfall classification. The goal was to build end-to-end data pipelines and share key insights.

Data Collection & Wrangling

- Collected data from sources including CSV files, online APIs (yfinance), and HTML tables. Used Pandas for cleaning, handling missing values, and structuring data.

EDA & Visualization Methodology

- Used matplotlib and seaborn for visualizing distributions, correlations, and trends. Applied group-by operations and filtering for detailed insights.

Predictive Analysis Methodology

- Implemented regression and classification models including Ridge regression and logistic regression. Applied pipelines, polynomial features, and grid search.

EDA Visualization Results

- Displayed housing price distributions, heatmaps for correlation, pie charts for categorical variables (e.g., diabetic vs. non-diabetic). Identified feature importance visually.

SQL Analysis Results

- Connected to SQL database and used queries to explore datasets. Extracted key statistics and performed joins to gain deeper insights into relational datasets.

Folium Interactive Map

- Created a choropleth map showing office locations and sales distribution during recession using GeoJSON and Folium.

Plotly Dash Dashboard

- Built an interactive dashboard with dropdowns, filters, and plots using Plotly Dash. Enabled exploratory analysis for non-technical users.

Predictive Analysis Results

- Ridge Regression R^2: 0.648
- Polynomial Ridge Regression R^2: 0.700
- Rainfall classifier evaluated with precision, recall, F1-score and confusion matrix.

Conclusion

- The capstone integrated various data science techniques. The predictive models and dashboards created provide actionable insights and serve as strong portfolio projects.

Creativity & Innovation

- Added advanced modeling, interactive visualizations, and integration of external data sources to enrich analysis and storytelling.