**Data Analysis and Machine Learning Project**

I completed an exciting data analysis and machine learning project that involved extracting insights from a dataset of employee information using Python , popular data science libraries and Power BI. Here's a breakdown of what I accomplished and the tools I used:

1-**Data Cleaning and Preprocessing:**

* Leveraged pandas for efficient data manipulation, including handling missing values, standardizing columns, and removing duplicates.

2-**Statistical Analysis:**

* Calculated descriptive statistics (mean, mode, sum) to answer specific 28 questions about the dataset, such as average income, percentage of employees meeting certain criteria, and identifying common attributes.

3-**Data Visualization:**

* Various plots are created using seaborn and matplotlib to visualize distributions (**sns.boxplot()**), relationships (**plt.scatter()**), and correlations (**sns.heatmap()**).
* Plotting the mean income by the number of cars owned using **plt.bar()**.

4-**Machine Learning Model Training:**

* + Features are selected using **SelectKBest** and **f\_regression** to pick the top features based on F-test scores.
  + Label encoding categorical variables (**Gender**, **Marital Status**, **Education**, **Region**) to numeric format.
  + Splits the data into training and testing sets (**train\_test\_split()**).
  + Implements GridSearchCV with Support Vector Regression (SVR) to find the best hyperparameters for the SVR model.
  + Trains a SVR model and evaluates its performance using RMSE on the test set.
  + Similarly, trains a Linear Regression model and evaluates its performance.
  + Similarly, DecisionTree Regressor model and evaluates its performance.

**5-Model Evaluation and Result Visualization:**

* + Plots the predicted vs. true values using matplotlib to visually inspect the model predictions.

6-interactive Power BI Dashboard

Analyze employees data