ROC COMPANY ANALYSIS

PHASE 3:DEVELOPMENT PART 1

SUBMITTED BY:

VELPUCHARALA VENKATA SUBBA REDDY

au723921243054

[subbar262@gmail.com](mailto:subbar262@gmail.com)

Q: START BUILDING AI DRIVEN EXPLORATION AND PREDICTION PROJECT BY LOADING AND PREPROCESSING THE DATASET

1.Load The Data.

**11111jhfug**

import pandas as pd

data = pd.read\_csv("dataset.csv")

print(data.head())

**2.Data Preprocessing**:

Data preprocessing is essential for cleaning, transforming, and preparing the data for machine learning. Here are some common preprocessing steps:

**3.Handle Missing Values**:

You can check for missing values and decide how to handle them. You might choose to remove rows with missing values or impute them with the mean, median, or a custom

Check for missing values

print(data.isnull().sum())

Handle missing values (e.g., fill with mean)

data.fillna(data.mean(), inplace=True)

**4.**Create new features or modify existing ones to improve model performance.

**5.Encode Categorical Variables**:

If your dataset contains categorical variables, you should encode them into numerical values. One-hot encoding is a common technique

data = pd.get\_dummies(data, columns=['categorical\_column'])

**6.Split the Data into Training and Testing Sets**:

Split the dataset into a training set and a testing set. This allows you to evaluate your model's performance.

from sklearn.model\_selection import train\_test\_split

X = data.drop("target\_column", axis=1)

y = data["target\_column"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**1.Data Exploration**:

Perform exploratory data analysis (EDA) to understand your data. You can create visualizations and summary statistics to identify patterns and relationships.

**2.Build and Train Machine Learning Models**:

You can now use various machine learning algorithms to create predictive models. For example, you can use Scikit-Learn to build regression, classification, or clustering models.

**3.Evaluate and Fine-Tune Models**:

Assess model performance using appropriate metrics and make adjustments as necessary. This might involve hyperparameter tuning or selecting different algorithms.

**4.Deployment and Prediction**:

Once you have a well-performing model, you can deploy it to make predictions on new data.

**1.Load the Data**:

Load the company registration dataset using Pandas:

Load the dataset

data = pd.read\_csv("company\_dataset.csv")

Display the first few rows to understand the data

print(data.head())

**2.Data Preprocessing**:

let's perform some common data preprocessing tasks:

* **Handle Missing Values**:

Check for missing values and decide how to handle them, based on your data and analysis requirements.

print(data.isnull().sum())

# Handle missing values (e.g., fill with mean, median, or custom value)

data.fillna(data.mean(), inplace=True)

* **Feature Engineering**:

Create new features or modify existing ones if it helps with your analysis. For instance, you might want to calculate new metrics based on the existing columns.

* **Encode Categorical Variables**:

If your dataset contains categorical variables, encode them into numerical values using one-hot encoding or label encoding, depending on the context.

data = pd.get\_dummies(data, columns=['categorical\_column'])

**3.Scaling or Normalization**:

Depending on the analysis or machine learning algorithms you plan to use, you might need to scale or normalize numerical features

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

data[['numerical\_column1', 'numerical\_column2']] = scaler.fit\_transform(data[['numerical\_column1', 'numerical\_column2']])

**4.Data Exploration**:

Perform exploratory data analysis (EDA) to understand your data. This may involve creating visualizations, summary statistics, and identifying patterns and relationships.

**5.Carry Out Your Analysis**:

At this point, you can proceed with the specific analysis you have in mind. You might use statistical techniques, machine learning models, or other methods based on the nature of the dataset and your objectives.

**6.Data Preprocessing**:

Data preprocessing is essential for cleaning, transforming, and preparing the data for machine learning. Here are some common preprocessing steps:

**7.Handle Missing Values**:

You can check for missing values and decide how to handle them. You might choose to remove rows with missing values or impute them with the mean, median, or a custom value.

Check for missing values

print(data.isnull().sum())

Handle missing values (e.g., fill with mean)

data.fillna(data.mean(), inplace=True)

**Split the Data into Training and Testing Sets**:

Split the dataset into a training set and a testing set. This allows you to evaluate your model's performance.

from sklearn.model\_selection import train\_test\_split

X = data.drop("target\_column", axis=1)

y = data["target\_column"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

1. **Data Exploration**:

Perform exploratory data analysis (EDA) to understand your data. Create visualizations and summary statistics to identify patterns and relationships.

**2.Build and Train Models**:

Depending on your specific analysis goals, you can apply various machine learning algorithms or statistical techniques.

**3.Evaluate and Fine-Tune Models**:

Assess model performance using appropriate metrics and make adjustments as necessary. This might involve hyperparameter tuning or selecting different algorithms.

**4.Deployment and Prediction**:

If applicable, deploy the model to make predictions on new data.

Top of Form