FAST LANE CAREER

Data Analysis Projects

Dear Intern,

We're excited to share three flexible projects with you. Complete any two to be eligible for a Certificate of Completion and Internship Completion Certificate. Submit details via the Submission Form.

Project Details:

- Project 1 : "**Traffic Flow Prediction for Smart Cities**"
- Project 2 : "**Stock Price Analysis**"
- Project 3 : "**Healthcare Data Analysis**"

Project 1: Traffic Flow Prediction for Smart Cities

- **Objective:** Predict traffic flow patterns for optimal transportation.
- **Dataset:** Metro Interstate Traffic Volume from UCI ML Repository.
- **Goals:**
  - **Data Collection and Exploration:**
    - Collect historical traffic data from sensors, cameras, etc.
    - Explore the dataset for structure, variables, and patterns.
  - **Data Preprocessing:**
    - Handle missing data, outliers.
    - Convert categorical variables to numerical formats if needed.
    - Normalize or scale numerical features.

- **Feature Engineering:**
  - Extract relevant features (time, day, weather).
  - Create lag features to capture historical traffic patterns.
- **Exploratory Data Analysis (EDA):**
  - Visualize traffic data to identify trends, patterns, correlations.
  - Analyze traffic flow at different times and days.
- **Conclusion of the analysis.**

Project 2: Stock Price Analysis

Source - https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm

GITHUb - https://github.com/sajal2692/data-science-portfolio/blob/master/Stock%20Market%20Analysis%20for%20Tech%20Stocks.ipynb

- **Goals:**
  - Discover and explore stock market data.
  - Use yfinance/python library for NSE.
  - Visualize stock information using Seaborn and Matplotlib.
  - Analyze risk based on previous performance.
  - Predict future stock prices using LSTM.
- **Project Goals:**
  - Change in stock price over time.
  - Daily return of the stock on average.
  - Moving average of various stocks.
  - Correlation between different stocks.

- Value at risk by investing in a particular stock.
- Attempt to predict future stock behavior (e.g., Predicting the closing price of Apple Inc. using LSTM).

Project 3: Healthcare Data Analysis for Diabetes Prediction

- **Objective:** Develop a model to predict diabetes likelihood based on patient health metrics.
- **Dataset: Pima Indians Diabetes Database from Kaggle.**
- **Project Goals:**
  - **Data Collection:**
    - Collect the Pima Indians Diabetes Database or a relevant diabetes dataset.
    - Ensure compliance with data protection and privacy regulations.
  - **Data Exploration:**
    - Explore the dataset to understand feature distribution and diabetes outcome.
    - Identify missing values or outliers.
  - **Data Preprocessing:**
    - Handle missing data through imputation or removal.
    - Normalize or standardize numerical features.
    - Encode categorical variables into numerical formats if necessary.
  - **Exploratory Data Analysis (EDA):**
    - Visualize feature distribution and relationships with diabetes outcome.
    - Explore correlations between different features.
    - Identify potential patterns or trends.
  - **Feature Selection:**

        - Evaluate feature importance for diabetes prediction.
        - Consider techniques like correlation analysis or feature importance from ML models.
    - Conclusion.

We are excited about the prospect of working with you and confident that your skills will be a valuable asset to our team.

Sincerely,

Fast Lane Career