Crypto Risk and Return Analysis Project Documentation

# Introduction

This project focuses on analyzing minute-level cryptocurrency data to assess risk and return patterns. Using PySpark for scalable data processing and Plotly for interactive visualizations, the project extracts key insights such as hourly price trends, volatility patterns, and average close prices.

# Technologies Used

- PySpark: For distributed data processing and transformation.  
- SparkSession: Entry point to interact with Spark functionality.  
- Plotly: For creating interactive and high-quality visualizations.  
- Python: Core programming language used for scripting and analysis.

# Project Setup

1. Start a Spark session using SparkSession.  
2. Load the dataset from a CSV file using spark.read.csv with header and schema inference.  
3. Transform the data by converting timestamps, extracting hour and date, and calculating volatility.

# Analysis Steps

## 1. Average Close Price per Hour

Group the data by hour and compute the average close price to understand hourly price behavior.

## 2. Maximum Volatility per Day

Calculate the maximum volatility (high - low) for each day to identify the most volatile trading days.

## 3. Top 5 Most Volatile Trading Days

Sort the daily volatility values and extract the top 5 days with the highest volatility.

## 4. Hourly Price Trend for a Selected Date

Filter the dataset for a specific date and analyze the hourly average close price to observe intraday trends.

# Dataset Structure

The dataset contains the following columns:  
- open  
- high  
- low  
- close  
- volume  
- marketCap  
- timestamp  
- crypto\_name  
- date

# Conclusion

This project demonstrates how PySpark can be used to efficiently process large-scale cryptocurrency data and extract meaningful insights. The use of Plotly enhances the interpretability of results through visualizations. The analysis provides a foundation for understanding crypto market behavior and identifying high-risk periods.