Data Collection and Preprocessing Phase

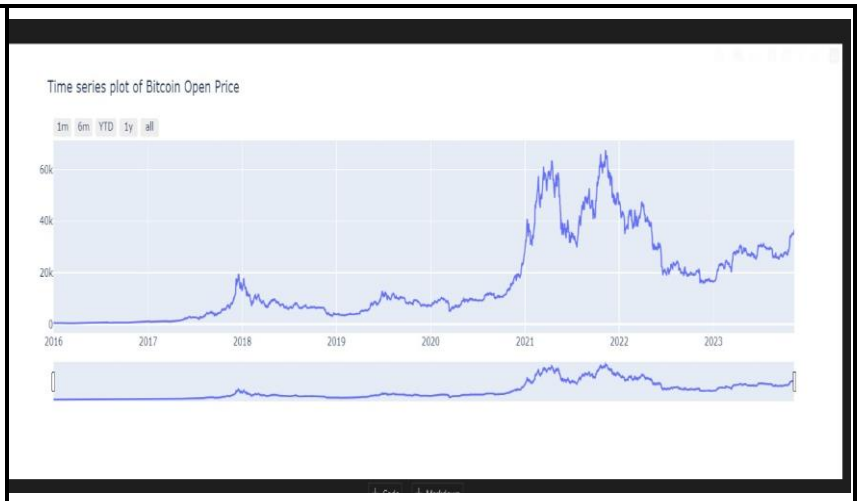
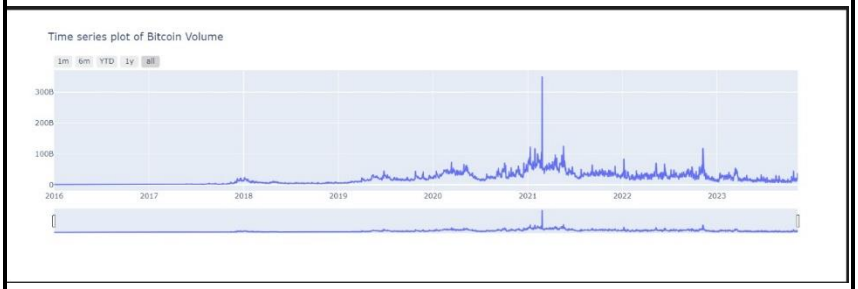
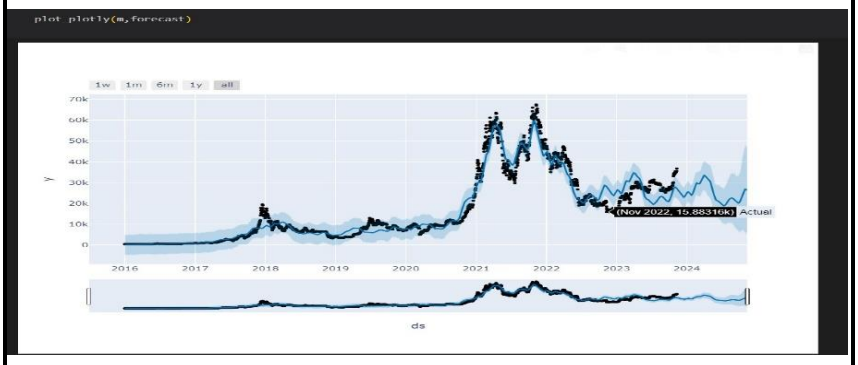
Date	03 October 2024
Team ID	LTVIP2024TMID24963
Project Title	Time Series Analysis For Bitcoin Price Prediction Using Prophet
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	Dimension: 2871 rows × 6 columns Descriptive statistics: `df.describe()`

	Open	High	Low	Close	Adj Close	Volume
count	$2,871.00	$2,871.00	$2,871.00	$2,871.00	$2,871.00	$2,871.00
mean	$16,457.17	$16,846.61	$16,036.22	$16,468.45	$16,468.45	$19,169,321,501.60
std	$16,146.91	$16,543.08	$15,698.06	$16,145.49	$16,145.49	$19,431,025,984.72
min	$365.07	$374.95	$354.91	$364.33	$364.33	$28,514,000.00
25%	$4,056.91	$4,117.19	$3,973.48	$4,067.15	$4,067.15	$3,675,985,000.00
50%	$9,522.33	$9,698.23	$9,305.91	$9,522.98	$9,522.98	$15,656,371,534.00
75%	$26,678.13	$27,073.72	$26,326.61	$26,736.56	$26,736.56	$29,901,790,331.00
max	$67,549.73	$68,789.62	$66,382.06	$67,566.83	$67,566.83	$350,967,941,479.00

Visual Analysis	   

Outliers and Anomalies	-
Data Preprocessing Code Screenshots	
Libraries imported	

```
import pandas as pd
import numpy as np
import yfinance as yf
from datetime import datetime
from datetime import timedelta
import plotly.graph_objects as go
from fbprophet import Prophet
from fbprophet.plot import plot_plotly, plot_components_plotly
import warnings
warnings.filterwarnings('ignore')
pd.options.display.float_format = '${:,.2f}'.format
```

Data Transformation

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 2871 entries, 2016-01-01 to 2023-11-10
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Open       2871 non-null   float64
 1   High       2871 non-null   float64
 2   Low        2871 non-null   float64
 3   Close      2871 non-null   float64
 4   Adj Close  2871 non-null   float64
 5   Volume     2871 non-null   int64
dtypes: float64(5), int64(1)
memory usage: 157.0 KB
```

```
df.isnull().any()
```

```
Open         False
High         False
Low          False
Close        False
Adj Close    False
Volume       False
dtype: bool
```

```
df1=df[['Date','Open']]
df1.head()
```

	Date	Open
0	2016-01-01	$430.72
1	2016-01-02	$434.62
2	2016-01-03	$433.58
3	2016-01-04	$430.06
4	2016-01-05	$433.07

```
newn={
    "Date":"ds",
    "Open":"y",
}
df1.rename(columns=newn,inplace=True)
df1.head()
```

	ds	y
0	2016-01-01	$430.72
1	2016-01-02	$434.62
2	2016-01-03	$433.58
3	2016-01-04	$430.06
4	2016-01-05	$433.07

Feature Engineering	Attached the codes in final submission.
Save Processed Data	

```
import pickle
pickle.dump(m,open('fbcrypto.pkl','wb'))
```