Project 2 - Bit Coin Prices - Part 4 - Perform Data Analysis

Dataset Link: https://www.kaggle.com/datasets/chakradharmattapalli/bitcoin-prices

Importing Necessary Libraries

```python
import numpy as np
import pandas as pd
```

Import and read dataset

```python
#Write Your Code Here
df=pd.read_csv('bitcoin_prices2.csv')
df.head()
```

```
   open  high   low  close  tick_volume  year  month  day
0  5.26  5.47  4.80   5.21        69150  2012      1    2
```

```
1  5.22  5.29  4.65  4.88      125170  2012    1    3
2  4.88  5.70  4.75  5.57      131170  2012    1    4
3  5.57  7.22  5.57  6.94      182328  2012    1    5
4  6.95  7.21  6.13  6.70      218077  2012    1    6
```

Data Cleaning

a. Missing Value

```
#Write Your Code Here
df.isnull().sum()
```

```
open            0
high            0
low             0
close           0
tick_volume     0
year            0
month           0
day             0
dtype: int64
```

b. Duplicate data

```
#Write Your Code Here
df.duplicated().sum()
```

```
0
```

Data Analysis

1.   What was the average closing price of the stock for each month in the year?

```
#write your code here
# convert columns to datetime format and assign to new 'date' column
df['date'] = pd.to_datetime(df[['year', 'month', 'day']])

# group data by month and calculate average closing price for each
month
avg_close_price = df.groupby(df['date'].dt.strftime('%Y-%m'))
['close'].mean()

# print resulting series
avg_close_price
```

```
date
2012-01      6.051818
2012-02      5.100952
2012-03      4.916364
2012-04      5.002857
2012-05      5.078261
             ...
2020-08   11631.657143
```

```
2020-09    10684.284545
2020-10    11901.793636
2020-11    16650.635238
2020-12    21788.898182
Name: close, Length: 108, dtype: float64
```

1.   What was the average daily price range (high - low) for a given month and year?

```python
#write your code here
year = 2020
month = 1

# filter data by year and month
data_filtered = df[(df['year'] == year) & (df['month'] == month)]

# calculate daily price range and store it in new 'price_range' column
data_filtered['price_range'] = data_filtered['high'] -
data_filtered['low']

# calculate average daily price range for selected month and year
avg_price_range = data_filtered['price_range'].mean()

# print resulting average daily price range
print(f'The average daily price range for {month}/{year} was:
{avg_price_range}')
```

```
The average daily price range for 1/2020 was: 406.90636363636366

C:\Users\HP\AppData\Local\Temp\ipykernel_3400\1296620740.py:9:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  data_filtered['price_range'] = data_filtered['high'] -
data_filtered['low']
```

1.   What was the total tick volume for each year in the dataset?

```python
#Write your code here
yearly_tick_volume = df.groupby('year')['tick_volume'].sum()
yearly_tick_volume
```

```
year
2012     16069420
2013    344894281
2014      2400003
2015      2384980
2016      4923692
2017    339002469
```

```
2018    558906751
2019    415228459
2020    478490502
Name: tick_volume, dtype: int64
```