

Technolution

Pre placement task report

Submitted by:

Mahesh Sasikumar

Data Analytics

IIITMK

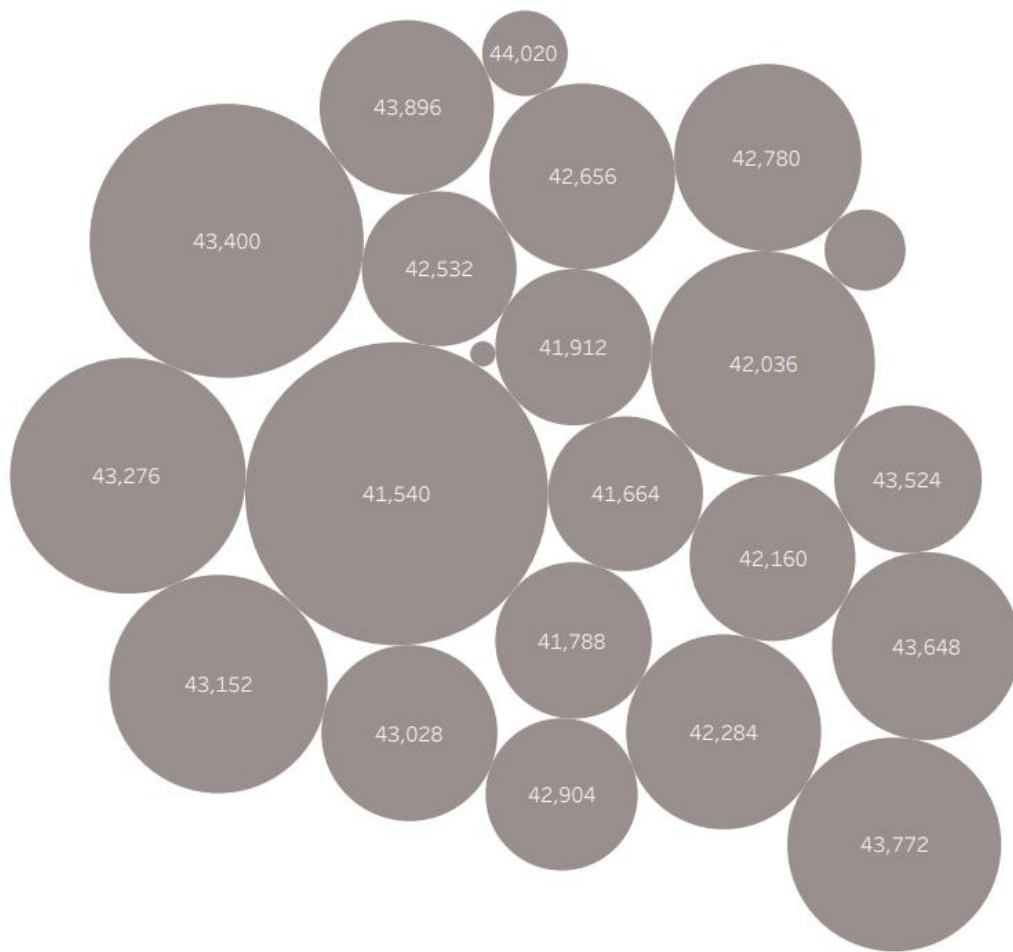
Question

The task was to create a pipeline with python, sql ,tableau and to create a dataframe with tweetcount, timestamp, real-time bitcoin price and to visualize and analyse the data that is created

Solution

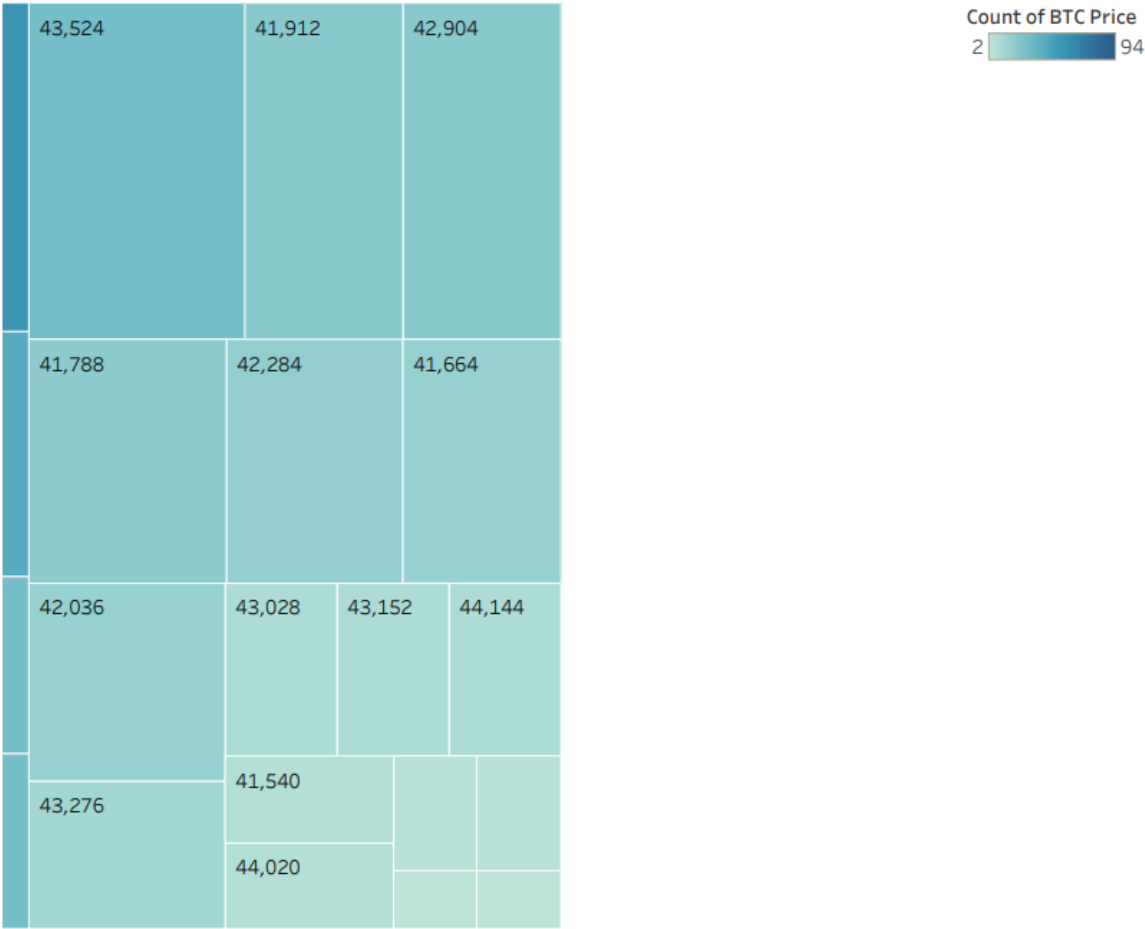
- The tweets were extracted from twitter using twitter api
- The timestamp of the tweets about bitcoin was extracted from the api
- But while coming to taking the count of tweets for three day which is 72 hours within a timespan of 15 min needs high amount of computation power and therefore there is a keyboard interrupt for the same.
- Even though use of datastructures such as dict,hash-table might help but it might also will not sufficient as the data-generated is 200k per day. That is about 200k tweets per day
- Therefore Bitcoin price along with timestamp is converted as a dataframe
- The dataframe is connected to sql and visualized using tableau

Visualization using Tableau



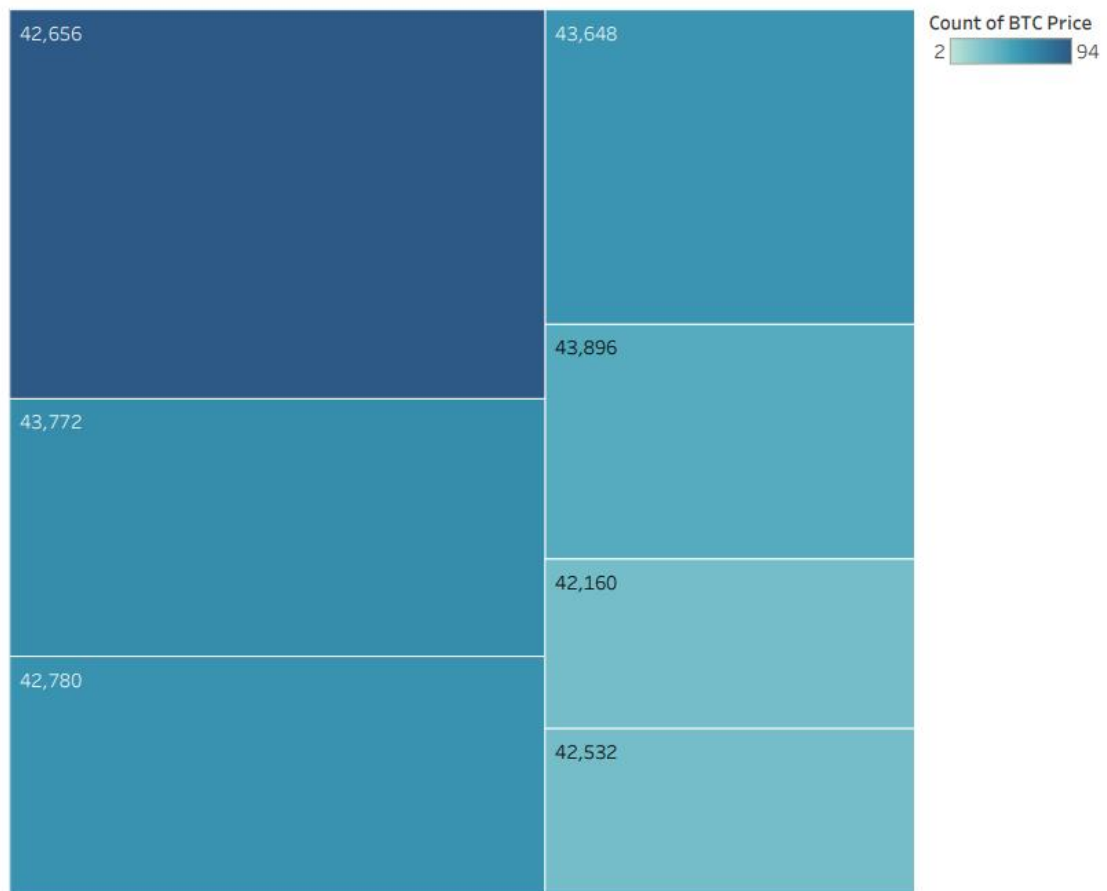
BTC Price (bin). Size shows variance of BTC Price. The marks are labeled by BTC Price (bin).

Sheet 1



BTC Price (bin). Color shows count of BTC Price. Size shows count of BTC Price. The marks are labeled by BTC Price (bin).

Sheet 1



BTC Price (bin). Color shows count of BTC Price. Size shows count of BTC Price. The marks are labeled by BTC Price (bin).