



**Capstone Project
EDA
(Credit Card Fraud Detection System)**

Devashish Kapadia

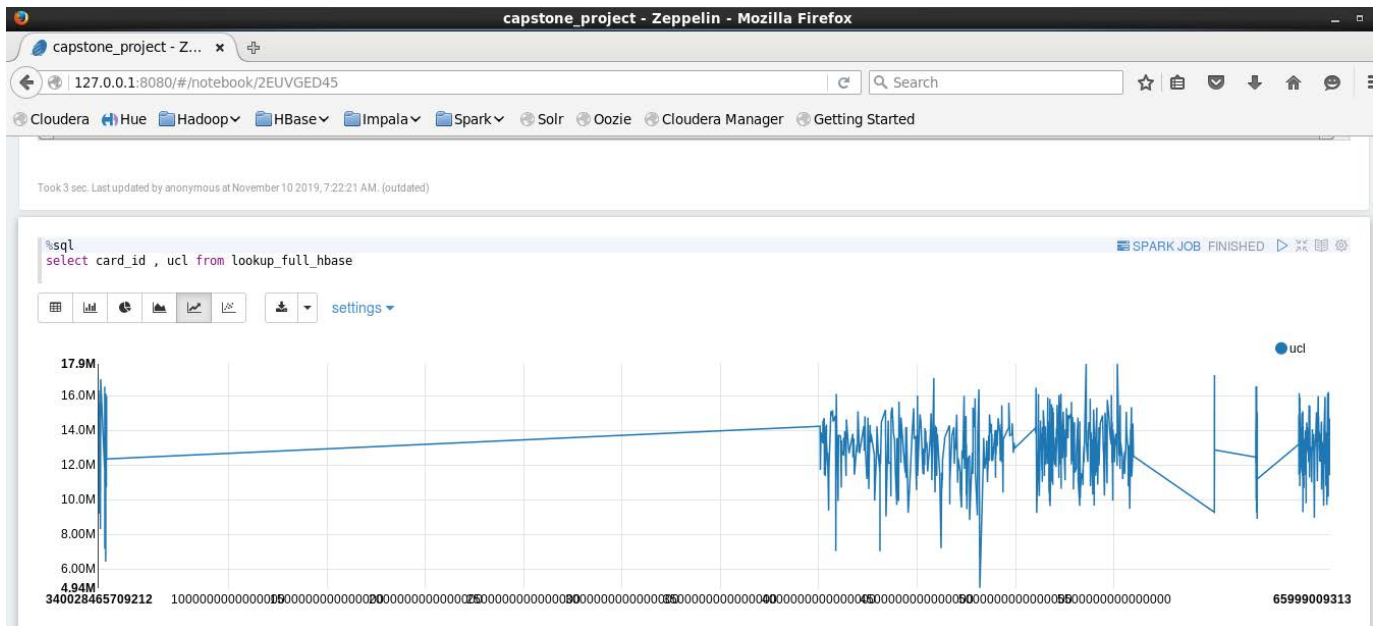
Content

- UCL ON LINE CHART
- TOP 5 ZIPCODES WITH MAXIMUM UCL VALUE
- PLOT THE CREDIT CARD 'SCORE' WITH THE UCL VALUE. CALCULATE THE CORRELATION BETWEEN THEM

Get a fair idea about the UCL values and plot it over a line chart.

--Query to plot UCL on line chart with respect to card_ID

```
SELECT CARD_ID , UCL FROM LOOKUP_HBASE_FULL;
```



Identify the top 5 Zipcodes with maximum UCL value.

-- Query used to get the value

```
SELECT CARD_ID , UCL , POST_CODE FROM LOOKUP_FULL_HBASE ORDER BY UCL DESC
```

```
LIMIT 5;
```

The figure is a screenshot of a Zeppelin notebook interface showing the results of a SQL query. The query is: '%sql select card_id , ucl , post_code from lookup_full_hbase order by ucl desc limit 5'. The results are displayed in a table with three columns: 'card_id', 'ucl', and 'post_code'. The table contains five rows of data. The status bar at the top right of the table area indicates 'SPARK JOB FINISHED'.

card_id	ucl	post_code
5517690092508277	17878791	10927
5359301116643451	17857012	52241
6011780257723719	17211782	80726
4585242021970523	17051056	25849
349996104851677	16976910	17876

Plot the credit card 'score' with the UCL values (normalize the values if required, for better visualization) and calculate the correlation between them.

--Query to plot

SELECT UCL, SCORE FROM LOOKUP_FULL_HBASE ;

NOTE:- To normalize plotted value of ucl on the Y axis of the scatter chart

--Query to calculate the correlation

SELECT CORR(SCORE,UCL) FROM LOOKUP_FULL_HBASE;

ANS:- **-0.006289702736715526**

