DATA ANALYST CHALLENGE

Tamara Martínez

Challenge #1 – Data Vision📊

This is a problem that, unfortunately, a lot of companies encounter, despite their size, purpose, or whatever they do. Data is essential for decision making and target creation, hence the importance of its accuracy and also availability.

The organization and scheme of the modern data stack of a company can make the process of decision making, insights discovery, OS, data sharing, etc. a whole lot easier or even more complicated. Knowing that, I would propose to focus for now in one part of the data stack an improve it, specifically in the loading part, so all data can be integrated in a data warehouse, and consequently data analysts, data scientists, software engineers, or anyone can transform this data for specific matters.

Nowadays, there are a lot of platforms and software that offer various solutions. Therefore, by getting one that offers ETL automation will make all this processes more efficient, so that the company can focus in what's important. For example, Fivetran, a company that makes centralizing data from any source to any data warehouse effortless. It's a fully automated ETL tool that focuses more on extraction and loading. Fivetran automates data integration from source to destination, providing data that a team can analyze immediately.

So, why keep doing this manually when automation eliminates the need for manual coding and provides a convenient and easy way to manage data flows and implement basic data transformations.

Another idea would be to create various dashboards targeted to specific areas and questions, and have them in one platform so that anyone could see them and extract, via the charts and other elements of these, the information more frequently needed. In case someone requires to solve a specific query or create and view a particular chart it can be done by themselves or even ask someone in the analysis or IT department. A platform that can do this and so much more is Looker by Google, but there are plenty more out there, it's just a matter of finding the best fit for the company.

Challenge #2 – The tools 🔑

View attached jupyter notebook.

Challenge #3 - The Investor's questions

- ¿How many clients does Kueski have?

  SELECT COUNT (user_id)

  FROM users;

- ¿What is the share of clients by state?

  SELECT COUNT (user_id), state

  FROM users

  GROUP BY state;

- ¿Which state is the one with the greatest amount of loans?

  SELECT state, COUNT (state)

  FROM users GROUP BY state

  HAVING COUNT (state)= (

  SELECT MAX (mycount)

  FROM (

  SELECT state, COUNT (state) mycount

  FROM users

  GROUP BY state));

- ¿Which is the user who has the greatest amount of interest paid?

  SELECT user_id, MAX (interest_rate)

  FROM loans

  ORDER BY user_id;

- ¿What is our average loan amount?

  SELECT AVG (loan_amount)

  FROM loans;

- ¿What is our loss rate segmented by the disbursed month? Loss rate = principalpaid / loanamount

  SELECT

```sql
TO_CHAR (disbursed_date,'YYYY-MM') as month,
(principal_paid/loan_amount) as loss rate
FROM loans
GROUP BY months;
```