Octopus Energy Data Analyst Case Study

You've been provided with a sample database of calls in and out of Octopus Energy. You've been asked a series of questions by the head of operations who wants to find out more about the call activity.

Some guidelines:

- Please present your queries, working, answers to the first 2 sections in a single jupyter notebook. Please send an html version of your notebook and a PowerPoint presentation or any other support you prefer to highlight the elements for presentation and discussion with the Head of operation.

- You can present your answers to section 3 in whichever tool you choose

- Please don't spend too much time on this. The aim is to get as far as you can in 3-4 hours of work. If you can't finish, feel free to add notes on what you would have done.

- You don't have much time so if you get stuck on a question, move on

- If anything isn't clear feel free to email me (julie.deprieck@octopusenergy.fr) to clarify

Data: You've been provided with a sqlite database with 3 tables:

- **account** - this is one row per account and contains data on each customer account included in the call table

- **call** - this is a table with one row for each phone call our operations team took or made. It contains useful data about each call. Talk time is in seconds. Called_at is the datetime the call started in UTC.

- **call_reason** - this is a look up table for the call reason_id

Section 1: SQL Only

For this section please do all of your working in SQL

Question 1: How many calls were made on 13th March 2017?

Question 2: Which sales channels had the top 3 most calls in April 2019 and how many calls did they each have?

Question 3: Which agents (ids) have taken over 12 calls in a day?

Question 4: Which hour of the day typically has calls with the longest talk time?

Question 5: Which agent took the first call with the category Iresa Enquiry ?

Question 6: What's the average time-on-supply of customers when they call for the first time? Time on supply of a customer is defined as call date minus sign up date.

Section 2: Python and SQL

For this section you can use Python and SQL. Feel free to illustrate your answers with simple visualisations where you feel appropriate. Use any Python libraries you like.

Question 7: Do call agents typically decrease their average call talk times the longer they work at Octopus?

Question 8: How does call talk time relate to the call category?

Question 9: On the 17th May 2019, what was the maximum number of agents on the phone **at the same time**? (bonus, plot a graph of number of occupied agents throughout the day)

Section 3: Simple dashboard

Using whatever dashboarding, BI, data app or visualisation library you like. Create a simple call dashboard for the ops management team.

Feel free to add in whatever plots and metrics you think are relevant and useful for the ops management team.

There is no correct answer, we are looking to see how you approach the problem.

Assume that today is the last day in the dataset.

Please send the dashboard in a format we can easily view without installing the software. This could be an html, pdf, image file or link to an online dashboard.