

Candidate Handout: Telco Case Study

Business problem:

A telecom company has approached QuantumBlack to help optimise their infrastructure maintenance strategy. Some issues can be fixed remotely, but others require an engineer to be deployed on site. The latter option is very resource intensive in terms of FTE days. You have arrived on the first week of the study and have started to map out and understand the data available.

This data exists as a natural by product of day to day operational systems and was not really designed with analytical use-cases in mind. You are one of the first Data engineers on-boarded on the study and would be helping with quick analysis and data exploration to start with. You have to use Python, PySpark or Spark to do this data analysis.

Brief about Data

- Alarms - Each tower has sensors to monitor and report issues. When sensor detects an issue, it generates an alarm.
 - Column ALARM_SRC contains three elements of the tower
 - If ALARM_SRC would be in JSON format. e.g. if it is `{"Network":"UK_TEL_LON","SubNetwork":"LON001","MeContext":"LON001LTE4813"}`
 - *Network*= shows the name of the network and it is a constant i.e. UK_TEL_LON
 - *SubNetwork*= shows the tower on which alarm is raised i.e. LON001
 - *MeContext*= shows the part of the tower having issue i.e. LON001LTE4813
 - Type of network - ALARM_SRC.MeContext also tells us if the network is 2G, 3G or 4G.
 - If MeContext contains **LTE** it is 4G network
 - If MeContext contains **UTM** it is 3G network
 - If MeContext contains **GSM** it is 2G network
- Ticket - If an alarm requires attention or some type of fix, a ticket is raised for Operations Team. Then the "Operation" team would either resolve it remotely or by visiting the site.
- Team - Details about the team who works on solving the ticket.

No.	Time (approx.mins)	Question
1	10	What might be the drivers for maintenance performance?
2	15	<p>After working with Data Scientists and industry experts, we've agreed to uncover the drivers of engineering team performance using a model. The first task is to assess the data available.</p> <ul style="list-style-type: none"> • What data quality issues would you be concerned about and check for? • Execute a test for completeness of data against the 'Issue type' field in the alarms.json file
3	10	<p>A simple initial definition of performance, is agreed to be the number of tickets a team resolves. To better understand the variances show:</p> <ul style="list-style-type: none"> ▪ team ▪ tickets_priority_1_count ▪ tickets_priority_2_count ▪ tickets_priority_3_count
4	15	<p>As input to the early model effort create a dataframe with:</p> <ul style="list-style-type: none"> ▪ year ▪ month ▪ team ▪ team size ▪ tickets solved in the current month ▪ tickets solved in the previous year
5	10	<p>A hypothesis is that the teams working on newer towers will take longer to resolve issues.</p> <p>A colleague began work on the below, but has left it unfinished...</p> <p>Create a function that accepts alarms records (or subparts) and returns the alarm network type.</p> <p>This is a chance to demonstrate your software engineering skills. Make sure this function is robust & ready for PR; consider testing & documentation.</p>
6	(Bonus question)	<p>Create a new version of the dataframe in Q4 with "count of tickets by network type".</p> <p>This should be aggregated by year, month, team. Use the above function to get the network type.</p>

