Lloyds Banking Group

Data Analyst Incubation

Sprint 2 : CRM Data Integration and Branch Analysis

Contents

Introduction

As a Data Professional with two Analytics projects in your portfolio, the team leader has recommended your skills to Simon, the Regional Operations Director. He explains that they have some Customer Relationship Management (CRM) datasets that have been siloed and the gap is preventing them from assessing Branch performance comprehensively. He has requested that the CRM data be integrated into the Quay Banking Database, ready for querying and analysis.

Business need

Regional Operations managers are currently struggling to draw meaningful insights regarding branch performance to inform their strategy. The business is keen to mitigate data silo-isation and better understand which branches are under-reporting and which are under-performing. They are also eager to discover branches that are performing well, to identify good practices that can be adopted elsewhere.

Email Received (high level requirements)

To: L.Hill@quaybanking.com
Subject: Request to integrate data and provide Branch Analysis.

Hello Lauren,

Following on from our meeting, we've decided that our division would really benefit from a Data Professional able to evaluate and load our siloed data into the Quay Banking Database. Following that, this person would then be ideal to query the data for insights to aid us in our strategy development. To do this, we'd really appreciate help in answering the following questions regarding branch performance:

1) Performance:
   a. Which 20% of branches are underperforming? – evident in "closed complaints without relief" and negative customer service feedback
      i. Can we rank the Call Centre Servers' performance according to Call duration and outcome ? (Advanced)
   b. Which are the top 5 branches per positive customer feedback?
      i. What are the characteristics of these branches that could contribute to this positivity ?Advanced)

2) Reporting CRM data:

     a. Which branches may be under-reporting their customer feedback? Highlight anything less than 2 standard deviations away from the mean
         i. Dive deeper into this pattern by separating branches in urban from those in more rural areas, as best you can. (Advanced)

3) Customer Complaints:
     a. How well are branches and customer service staff handling customer complaints?
         i. Which customer service staff are closing the most queries either as "Closed without relief" or "Closed with Explanation"?
        ii. Which branches are proportionally receiving the most complaints regarding "Account opening, closing, or management"?

Simon Moneybags Esq.
Regional Operations Director
Quay Banking Ltd

Brief from Project Manager

The Project Manager (Lauren) has scoped Simon's request further and determined that your skills and experience with multiple Analytics tech stacks make you well-suited to understanding the problem and framing it as an analytical challenge. The overall objectives are to:

- Evaluate and then integrate the CRM data into the Quay database

- Query the relevant data to answer the Regional Operations Director's questions in the email above

- Present your answers to Simon's questions using any tech stack, software, or tool that you judge fit for the task

- Record your approach and answers to Simons questions as a short video for your team

Additional resources:

The business analyst has provided you with 'QA Database Technical Documentation' of the existing database, including:

- `Logical Schema – Quay Banking`

- `Relationship mapping – Quay Banking`

The data tables you will be integrating into the Quay Bank database are:

`CRM_Call_Centre_Logs.csv`

`CRM_Events,csv`

`CRM_2015_17_reviews,csv`

Development brief from your supervisor

I would expect to see a few assets created during this project, so here is some guidance below on each stage:

Evaluation

1. (Using Excel, Tableau, PowerBI or Python +Pandas) Evaluate source data quality according to industry standards outlined by the British Computer Society (Professional Regulatory board):
   a. Completeness
   b. Consistency
   c. Uniqueness
   d. Accuracy
   e. Timeliness
   f. Validity
2. Write a short **Data Quality Summary** for each CRM file you have been given
3. Identify an appropriate Primary Key or determine how you will create one for each source
4. Plan how you are going to load the data into the existing database, paying special attention to data types and consistency (for example Client_ID would typically be consistently stored as a NVarchar in any table that references it)

Data Integration

1. Integrate the CRM tables into a unified, **coherent 'sandbox' version of the database** by loading the 3 CRM sources and relating them into a copy of the existing schema
2. Remember that a single unified database will prioritise data consistency and maintain referential integrity. You should set keys for the relationships, with constraints, and leverage the technical documentation that was used to construct the original database
3. Verify your database with queries and object explorer in SSMS; also, it is worth **documenting these verification steps**. Before promoting this new design to production, we would have to do some testing so it would be useful to get ahead of it
4. Create a **schema diagram** displaying the now extended database

NOTE – If you cannot relate one of the tables, move on and complete the others. Although it's not best practice, when the data is extracted for analysis at a later date, we can use a query to manually join the CRM tables to existing database entities. Do make a note of this in your writeup though and reflect the absence of a join in your own schema diagram.

Analysis

We want to prove the extended database design straightaway by answering Simon's questions from his initial email

1. Query the database to answer the Regional Director's questions, thinking about which fields from which entities are relevant. For example, when analysing branch performance, you will need to gather from multiple tables; Branch name, reviews, count(clients), dispute resolution.
2. You wont be able to see which Clients reside in which districts in the CRM Events table, but the Client table tells you which district every client is situated in and you can use that through a join
3. If you end up missing a column or two you felt is relevant but too difficult to extract, then make a note of this. Use this insight to make a recommendation for improvements in the next iteration of the database - we have been thinking about improving usability of the Quay Bank database anyway! Your ability to prioritise, identify areas for improvement will be useful.
4. Once you have your queries working, use any tech stack appropriate for analysing and visualising the answers to Simon's questions as an **Analytical Report**. If you find any of his questions ambiguous, document how you interpreted the question.
5. This is time sensitive so be transparent about what was possible and what could be improved later. *But please, do not write about how*

you ran out of time or what you would have done given another week. Do what you can, recommend what to develop further in the next iteration, then move onto the next task.

Communication

Record a **3-minute presentation video** to walk through your approach to the data integration challenge and your Analytical Report. Show how you integrated the data, extracted it and how each of his questions have been answered. Present this along with any visual analysis you created, ensuring the report can be viewed by Simon or myself. For example, if using Python for visualisations, do save any charts as jpegs or take screenshots and add these into to your Branch Analysis documentation. Power BI reports can be printed as pdfs.