CHALLENGES FACED

Datathon Team Name: AVON MAET

DataWarehouse: Snowflake DW

Database: DFA23RAWDATA

Schema: AVONMAET

In this comprehensive document, I aim to shed light on several noteworthy challenges that we encountered during the execution of our project. While these challenges may initially appear technical in nature, understanding and addressing them is fundamental to optimizing our future endeavors. I will elaborate on each of these challenges in a straightforward manner.

Challenge 1: Missing Data

One of the recurring issues we faced throughout the project was the prevalence of missing data points within both categorical and numerical variables. While we made concerted efforts to handle these null values effectively, it is imperative that we delve deeper to ascertain the root causes behind the missing data. Identifying the source of these gaps in our data is not merely a matter of data hygiene; it holds the key to establishing a robust and dependable data source for future projects. Consequently, a rigorous examination and documentation of data collection processes are essential.

Challenge 2: Timestamp Discrepancies

A noticeable inconsistency emerged concerning the formatting of timestamp records across various fact tables. While some timestamps were recorded in a comprehensive DateTime format (e.g., "2023-09-20 14:30:45"), others were captured in a truncated form (e.g., "2023-09-20 14:30"). This variance in timestamp formats could potentially hinder our ability to accurately correlate events in the project. Therefore, it is incumbent upon us to undertake a meticulous reconciliation process, harmonizing timestamp formats to ensure seamless temporal alignment.

Challenge 3: Lack of Robust Joining Keys

Our project encountered a significant challenge stemming from the absence of clearly defined joining keys for linking tables from diverse data sources. This dearth of well-defined joining keys necessitated the utilization of timestamp data as surrogate keys for table linkage. While this approach allowed us to progress, it introduced an additional layer of complexity and the potential for errors. In future projects, we should prioritize establishing robust joining keys to streamline data integration and minimize the risk of errors.

Challenge 4: Unfamiliarity with Snowflake Environment

A noteworthy challenge pertained to our team's limited familiarity with the Snowflake environment. While Snowflake is a powerful tool, our lack of expertise presented hurdles that impeded our project's efficiency. To address this challenge, it is imperative that we invest in training and skill development to harness the full potential of Snowflake and similar technologies in future endeavors.

Challenge 5: Absence of a Comprehensive Data Dictionary

Throughout the project, we grappled with the absence of a comprehensive data dictionary that would provide clear and concise definitions for tables and columns. This absence hindered our ability to understand and work with the data effectively. In future projects, we must prioritize the creation and maintenance of a well-documented data dictionary. This document should serve as a foundational reference, elucidating the meaning and context of each data element, facilitating clarity, consistency, and accessibility.

In summary, these challenges, although presented in a professional context, underscore fundamental considerations that are essential for successful project execution. Our commitment to addressing these challenges and incorporating the lessons learned into our future projects will undoubtedly lead to improved outcomes and enhanced data management practices.

Challenge 6: Creating the Dim_Location Table

In our project journey, we bumped into a tricky situation while trying to make the Dim_Location table – a vital part of our data setup. The trouble arose because we had location names that showed up more than once in our data. And here's the twist – those same locations had different longitude, latitude, and even belonged to different regions. Imagine trying to pinpoint where a place is when its name appears in multiple spots on the map, each with its unique set of coordinates and regional tags!

This made it a puzzle to create a Location Dimension table that made sense. We couldn't just lump all those location names together because they were quite different depending on their coordinates and regions. So, we had to put our thinking caps on to figure out how to organize this data in a way that would help us make sense of it all.