KPMG VIRTUAL EXPERIENCE PROGRAM BY FORAGE

1. Introduction

This is a virtual internship program case study with the company, KPMG. This program is hosted through the site Forage and enabled me to leverage my skills and tools as a Data Analyst in a real-world setting.

2. Internship Company

PricewaterhouseCoopers International Limited, commonly known as PwC, is a multinational professional services network of firms, operating under the PwC brand. It is one of the Big Four accounting firms, along with Deloitte, EY, and KPMG. With offices in 151 countries and more than 364,000 people, PwC is among the leading professional services networks in the world. They provide services to 87% of the Global Fortune 500 companies.

PwC firms offer services in Assurance, Tax, and Advisory, helping organizations and individuals create the value they are looking for. In FY23, PwC's gross revenues were US$53.1 billion. They are committed to building trust in society, solving important problems, and making progress on issues that matter from AI to climate change.

3. Scenario Company

Sprocket Central Pty Ltd, a medium-sized bike and cycling accessories organization, has approached a Partner in KPMG's Lighthouse & Innovation Team. Sprocket Central Pty Ltd. is keen to learn more about KPMG's expertise in its Analytics, Information, and Modeling team.

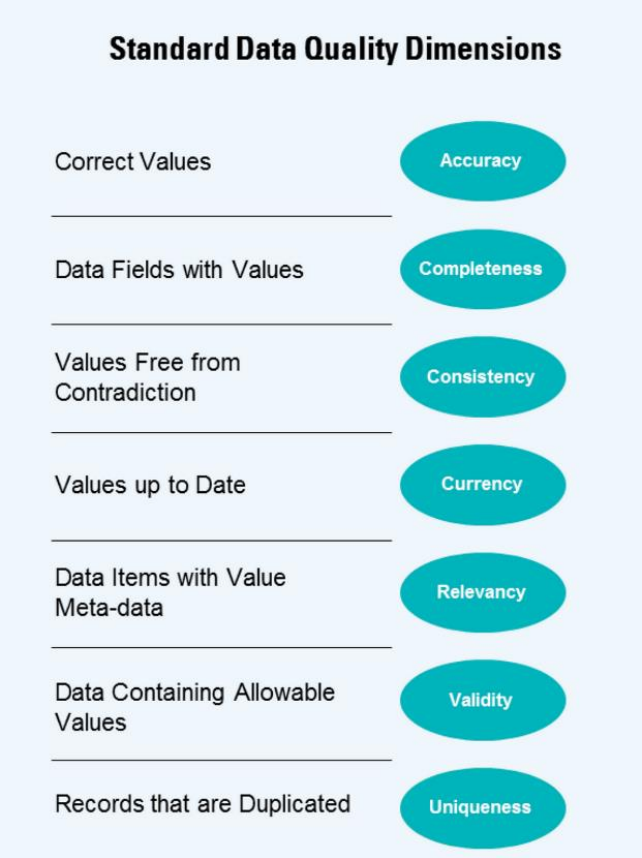
Primarily, Sprocket Central Pty Ltd. needs help with its customer and transaction data. The organization has a large dataset relating to its customers, but their team is unsure how to effectively analyse it to help optimize its marketing strategy.

The client provided KPMG with 3 datasets:

Customer Demographic
Customer Addresses
Transaction data in the past 3 months

4. Tasks

4.1 Task 1: Data Quality Check.



Data Quality Issues and Strategies to Mitigate These Issues.
I received the data sent earlier, and I have done the following data analysis on each dataset.

- The first thing I did was to make a copy of the data sets, and I subsequently went ahead to do some data cleaning on the duplicate.
- I used Microsoft Power Query in the data exploration and cleaning process. I imported the entire workbook into my power query.
- After exporting to Power Query, I discovered each Table/Dataset had multiple empty Colum, which I removed to reduce the size of the data.
- First Dataset/Table I analysed was the CUSTOMER ADDRESS TABLE. The table had the following data issues.

1. Accuracy: The Table had six Colum, I checked that each of the data are properly formatted, I changed the Colum's with number-to-number format, and The DOB was changed to a date format.
2. Completeness: The table had missing values, for missing values with a Text string I replaced them N/A.

3. Consistency: The Gender Colum had lots of inconsistency, it had lots of wrongly spelt words. The value with U was replaced as UNKNOWN. At the end I made sure we have just MALE, FEMALE, and Unidentified as our values.
4. The dob was changed from an INT sting to a date format. I discovered that the customer with customer id 34, has 1843 as DOB, I corrected the 1843 to 1943. There was so many missing dob, since we can't put N/A (which is a test string) and can't delete the empty rows. I simply put the Median date of all the date from the DOB Colum.
5. Validity The Tenure Colum had some empty rows, for easy calculation I replace empty rows with zero (0).
6. Relevancy: The default Colum was completely removed, I checked the Colum, and could not see its relevance to our data exploration and result.

The second Dataset/Table I analysed was the Customer Addresses. I use the promote row as header in power query to properly Title my Colum's. The Bike Related purchase Colum had missing value, it was replaced with zero (0). Other columns with missing values were replaced with N/A.

The Transaction table was also analysed, I converted the List Price and Standard price Colum to currency, and I chose USD as my currency. Missing values in Text format were replaced with N/A. The first purchase date Colum was in number format, I changed it to a date format.

Finally, Customer Demographic table was analysed, the missing values were replaced with N/A. the table had very little inconsistency.

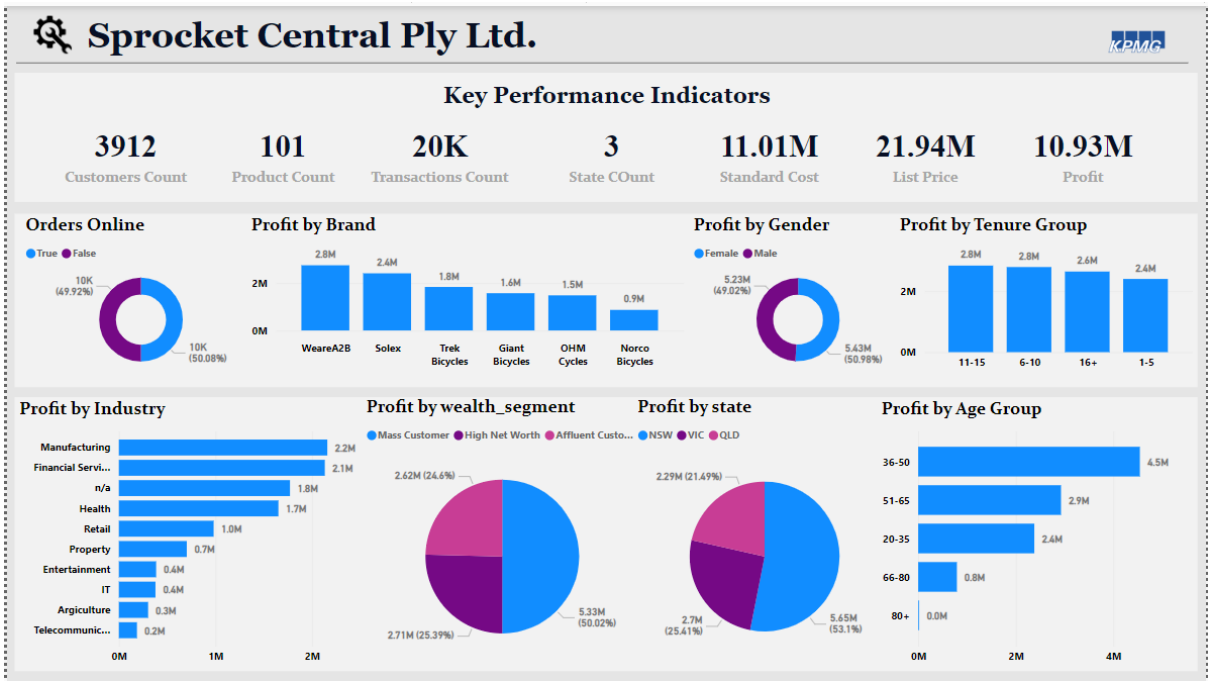
I imported my data back to Excel and I used to select all, Find/Select, Go to Special, Select blank Rows to validate my data and make sure no blank or missing values.

4.2 Task 2: Data Scope

Targeting high value customers based on customer demographics and attributes.

- In building this recommendation, we started with a PowerPoint presentation which outlined the approach which we will be taking.
- The client has agreed on a 3-week scope with the following 3 phases as follows - data exploration, model development and interpretation.
- Prepared a detailed approach for completing the analysis including activities – i.e. understanding the data distributions, feature engineering, data transformations, modelling, results interpretation, and reporting.

4.3 Task 3: Data Insights and Presentation



5. Conclusion

In an era where data rules the business world, honing data analytics skills is a smart career move. The KPMG virtual data analytics internship program offers a comprehensive and practical way to do just that. It equips aspiring data analysts with the knowledge, experience, and confidence needed to excel in this rapidly evolving field.