**Assessment Title**: Assignment For Senior BI Professional

Date: 11/01/2024

NAME: Pruthviraj Santosh Mohite

1. **Executive Summary**

• Overview

This PowerbI file is dedicated to the evaluation criteria for SR Analytics which is focusing on the sales analysis by the shopping platform of Shopify.

• Report Structure:

**Overall analysis contains following sections viz.**

Introduction

1. Sales For Existing and Acquired Customers
2. City wise Gross Sales Analysis
3. Average days between second and third purchase

**Data Sources**

• Data Connections:

1. Flat File (Excel file) – As per provided Excel files , data is introduced to the data model.

**Data Model (Star Schema)**

• Tables and Relationships:

Active relationships participating in the schema.

1. **Customer Details to Sales Details** (Many to many relation with both cross filter direction on Customer ID)

2. **Product Details to Sales Details** (Many to Many relation with both cross filter Direction on Product ID)

3. **Sales Details to Location Details** (Many to Many relation with both Cross filter Direction on location ID)

**Table Classification:**

Fact tables – **Sales Details**

Reason-

• No of Data rows are more

• Repeated values are observed

• The Observed Values can be aggregated

These are some of the properties of Fact table. Therefore I considered this table as fact table.

Dimension tables– **Location Details, Customer Details, Product Details**

**Reason-**

• Less data rows

• Number of repeated values are not seen

• Values cannot be aggregated

• The values from these tables can be used as a slicer or to filter the data.

Other tables – **Calendar Table.**

Reason- As a Data analyst I consider creating a common date table can lead to better data interpretation. This table is marked as a date table.

**Table Schema:** This data model uses **Star Schema** for the Model performance and Query Optimization.

**Star Schema** – Sales Details (Fact Table at the Center and Other dimension tables

• Data Transformations:

**Basic power query transformations are:**

1. Replacing Null values

2. Changing data Types

3. Promoted Headers

4. Removing Null Columns

5. Blank Country Name is replaced by other

**Visualizations**

• Key Visuals:

Key visuals used for the story telling are

1. Pie Chart

2. Bar chart

3. Column chart

4. Table and Matrix

5. Cards and multi row card.

6. Slicers

7. Cards

8. Textbox

**Interactive Elements:**

• visual level filters are used throughout the report.

1. To calculate the top 20 cities based on the total gross sales, visual level filters are used.

2. To show the number of new customers and existing customers visual level filter is used on the card.

**5. Calculations and Measures**

• Aggregate functions used for measures are

1. Sum

2. Divide

3. Count

4. Distinct count

**Filter functions**

1. Calculate

2. Filter

**Calculated Columns:**

• Customer tagging

• Order tagging

**Report Performance**

• Optimizations:

• Star Schema is used for better model performance and Optimization.

• Implicit measures are converted to explicit measures.

• Number of Visuals is limited in the report for minimum time of loading and slicer performance.

• Calender table is created explicitly for better performance of date time functions.

• Performance analyzer is used while creating the visuals to calculate the total time taken by the visuals in data loading.

**Considerations:**

• Data Quality

• Aggregations

• Data preview and cleanup

**Some Logics considered are**

1. As per the data customer lifecycle is given in the CUSTOMER\_LIFETIME\_DURATION column. I interpreted the lifecycle of customers as the date difference between the customer created date and last order date. Therefore the customers having the lifecycle of 0 are interpreted as new customers and others as existing customers

**How star schema will help in this report for scalability and Query Optimization**

1) Improved Query performance

The sales data table (Fact table) contains numeric values and other table contains descriptive values. We can aggregate this data based on different dimensions. This can lead to faster query performance.

2) Scalability

We can add multiple data sources and dimensions in the star schema by considering the primary and foreign keys.