Dear [Karem],

I hope this email finds you well. I would like to express our gratitude for sharing the three datasets from Sprocket Central Pty Ltd. Below; you will find a summary of the statistics derived from the data sets you provided. Please do not hesitate to reach out if any figures do not align with your expectations.

**Summary Statistics:**

Table Name: Opportunities, No. of Records =10000, No. of Fields = 30, Columns: [opportunity_id, order_number, flow_status_code, sold_customer_number, payment_type ,po_amount, po_approval_status ,brand ,model_year, model, Brand, Model, Engine Capacity, Body Type, Transmission Type ,Trim ,invoice_date ,invoice_amount ,attribute9 ,receipt_type ,receipt_numbe,r receipt_date, receipt_creation_date, receipt_amount ,move_numbe,r opportunity_status, wish_list ,so_creation_date, receipt_status, unit_list_price ]

**Data Quality Issues and Mitigations:**

**Important data not provided:** Sales, Profit, Revenue, ROI, ROM, approvals, and payments.

**Column names are not meaningful.**

**Missing Values:** we have 52% of empty data in opportunity_status column, I don't know what is the correct values of this data so if you can send me more details for higher work performance.

receipt_amount column also have 2% of null values but I replaced them with 0 if you want another value kindly contact with me.

model_year column has 1% of null values kindly if you can check data and tell me the correct values or I gonna drop it or replace it with not available.

**Inconsistent Values:** Inconsistencies in attributes, such as 'N', 'Y in wish_list column you have to send it to me with meaningful values I guess it mean no and yes but kindly approve it for me, receipt_stutes contain 'APP','REV'and 'UNAPP' provide me with the meanings for them, This problem appears again at opportunity_status column for 'WON' value, also  Brand column contain data that belongs to other null column not to brand itself. To prevent this issue in the future, we recommend implementing drop-down lists for user data entry, receipt_creation_date column contain data that I think may be invalid or not importanat.

**Duplicated values:** we have columns appeared twice and they contain same data so it gonna lead to confuse team and slow down our performance kindly provide me with more details about best column to use I consider to use the best at my point of view to them but it gonna be more accurate if you tell me exactly what to use.

**Inconsistent Data Types:** In some cases, the same attribute exhibits inconsistent data types (e.g., numeric values for some fields and strings for others) like Engine Capacity it is float but it has to be string type. We have converted selected records into numeric format and removed non-numeric characters from strings. To maintain data consistency, we suggest implementing data type constraints for fact tables in the database. Also attribute9 column it isn't representative name for a column and it's duplicated from payment_type but after checking I note that they aren't equal.

**Incorrect Date Values:** the dates aren't meaningful at all.

I will continue with data cleaning, standardization, and transformation as we progress with our model analysis. Questions may arise along the way, and all assumptions will be thoroughly documented. Once this process is complete, we would greatly appreciate the opportunity to liaise with your data Subject Matter Expert (SME) to ensure that all assumptions are aligned with Sprocket Central's understanding.

Please do not hesitate to contact me if you have any questions or require further clarification.

Kind regards,

Aya Tarek Ahmed