Hello [Stakeholders],  
  
I wanted to discuss some of the most important findings from our most recent examination of the user, product, and transaction datasets.

We found a number of issues with data quality during our study which are listed below that might affect how accurate the insights are.

**Key Data quality issues:**

* The Products dataset contains 215 duplicate rows. Before conducting any analysis, it is preferable to eliminate them from the dataset.
* The Products dataset contains 92% missing Category\_4 values. It indicates that since the majority of the data in the Category\_4 column are missing, it is not advised to perform any analysis utilizing it.

A graph with purple squares

AI-generated content may be incorrect.

* In User dataset, about 30% of the data in the Language column are missing, which makes it challenging to use the column for analysis.
* In User dataset, only about 5% of the data in other columns, such as Gender, State, and BirthDate, are missing. This is not a major concern but if any analysisis made based on these columns then their might be some inconsistency in the analysis outcome.
* Demographic analysis is additionally limited by the user dataset's gaps, which include thousands of users' missing gender, state, and birthdate information.
* Transaction Dataset has 25% missing data, which makes data analysis challenging.
* The histogram for Transaction dataset for Final Sales shows that a long tail extends to larger sums, while the majority of the final sale amounts are clustered around lower values. This suggests that the distribution is positively biased, with most transactions being smaller and a limited number of high-value sales.

A graph showing a graph of a sales graph

AI-generated content may be incorrect.

* Inconsistencies were also discovered in the quantity and sales fields, where non-numeric numbers—like "zero"—were entered in place of numerical values.

An important discovery from data analysis is that Tostitos with $ 260.99 in total sales is the top mark in the Dips & Salsa category. This suggests that the brand has effectively installed itself among customers and shows an important market preference for Tostitos products in the region. Many variables, including successful marketing campaigns, high -quality products, extensive accessibility and loyalty to consumer brand, may be responsible for high sales figures. Given this dominance, more consumer behaviour research will be useful. Knowing whether these transactions are the result of new or recurrent consumers can provide insight into trends in procurement and storage.

In addition, examining the seasonal patterns or promotional activities related to Toastito's sales can lead to significant insights into factors affecting the customer's purchasing decisions. Do consumers buy more Toastitos under athletic phenomena or holidays, for example in this year's special season? In that case, these trends can be used to increase sales further through strategic marketing initiatives. Contrary to the performance of Tostitos with other brands in the same category, you can reveal areas where other companies can increase the market share. Consumer demographic provision, the frequency of procurement and opportunities for product bunting can also help brands create strategies centered for successful competition. Future assessments will be more accurate if the data quality is strengthened, by ensuring that all transactions are accurately associated with brands.

In order to improve our analysis even more and to handle unresolved data quality problems, we need more help in some important areas. First, our study includes missing values ​​because the product data set lacks brand names. We need access to a comprehensive product mapping that connects each barcode with the respective brand to guarantee the correct conclusions. Integrating the external or updated product reference table will increase the purity of the data.

Second, the transaction data set contains non-numerical values ​​such as sales and quantity columns such as "zero" and missing items. Although some of these deviations are fixed, to ensure that we will help to examine the desired data format and business rules to ensure that we treat them properly. Does non-selections or lack of sales price suggest a problem with data capture, or should they be handled as a $0 transaction?

Access to transaction-related promotional or discount data would also be helpful in order to better understand consumer behaviour. Including this information could aid in the explanation of brand performance patterns if certain sales spikes are linked to promotional activities. In a similar vein, more detailed information on returning vs new clients in each category would offer insightful information about customer retention and brand loyalty.