Competencies

In this project, you will demonstrate your mastery of the following competencies:

* Apply descriptive analysis methods to address a business problem
* Interpret historical data to support areas for improvement
* Present visual solutions that provide value to end users and decision makers

Scenario

You are working as a business analyst for XYZ, a regional restaurant chain. Recently, your organization was acquired by ABC, another restaurant chain. The joint leadership team of the newly formed business needs to create a profitable three-year plan. They will look to your analysis to inform this plan, given that neither company has any analysis of its own. You will be the first analyst to look at the data of the newly formed business.

Both restaurant chains operate out of multiple locations, sell direct to consumers, and are fast-casual restaurants. However, they offer distinct products that cater to different customer bases. You need to provide a descriptive analysis of what the new joint customer base looks like and how that compares in terms of the demographics of the geographic regions they are now selling in, now that they are a single entity with locations in different geographical regions than their home base.

Your final product will be a suite of visualizations that explain to the joint leadership team the current state of this newly formed business, the customer profiles of the new organization, and the products that have historically sold best and worst and are the most and least profitable. Finally, you have been asked to compare the customer profiles to the regional demographics where the newly formed restaurant chain has locations. Using your analysis, the joint leadership team will craft a three-year plan to make the company profitable within the fast-casual industry segment, which is notorious for being competitive.

Directions

**Part One: Transforming Data and Identifying Trends**Using Power BI, clean and transform the data sets from XYZ and ABC to create a single data set that includes the aggregated data from both organizations to prepare it for analysis. You will then analyze the data to identify and explain patterns and trends and any additional findings.

1. **Using the XYZ and ABC customer data sets in the Supporting Materials section, identify errors and gaps in the data.**
   1. Which attributes do the data sets share? XYZ and ABC share the same attributes which includes but are not limited to: Customer Id, Order No, Item Description, Item Cost, Menu Price to mention but a few
   2. Identify whether or not data is either missing or incomplete. Explain how this impacts the integrity of the data. There were missing data in the Item Cost for company XYZ. The rows numbers for the two companies were not even meaning that one company had more information than the other. These Missing data can lead to inaccurate analysis. Also, incomplete data skewed results causing misinformed decisions.
   3. Which strategies did you employ to identify the errors and gaps in the data? Explain your reasoning. First, a manual inspection was carried out to review the samples in other to identify errors. The data was also filtered (conditional Formatting) to remove empty cells, error, duplicates etc. The methods applied ensured a thorough check for data quality issues, which allowed for a more reliable analysis and better decision making.
2. **Clean and transform the data.**
   1. Create new variables out of existing ones as needed so that you can address the joint leadership team’s questions and concerns. Explain any relationships or trends identified through your analysis.
      1. Analyze spending trends
         1. By month and season
         2. By gender
         3. By region
         4. By customers with and without children
      2. Analyze the best and worst selling products
         1. By region
         2. By month and season
         3. By gender
         4. By customers with and without children
      3. Analyze the profitability of products
         1. By region
         2. By month and season
         3. By gender
         4. By customers with and without children
3. **Create various visualizations for each variable.**
   1. Identify each variable in the data.
   2. For each variable, create appropriate visualizations and explain how they support the narrative.
      1. Analyze the shape of the visualizations.
      2. Is the data grouped in any particular way, or is it randomly scattered? Explain. The data was grouped by company though we had to merge it for better analysis. Some of the visuals are a representation of the individual companies while others represent both the individual and merged companies
      3. Are there any outliers? If so, what might they indicate?
   3. Identify and explain the patterns and trends in the customer base.
      1. What states have the highest and lowest profit margins?
      2. Which restaurant category has the highest and the lowest average check?
      3. Is there any pattern in average checks by time of day (AM or PM)? Yes! Usually lower average checks are found in the morning as customers often buy cheaper items like breakfast and coffee but from the Area chart we observe that company recorded high sales of $3,500 btw 6am and 9am in the morning and an average sale of $4,290 around lunch time
         1. What about by month of the year? The peaks were located in April, August and November
   4. Produce summary statistics for each variable.
      1. Include the following:
         1. Central tendency
         2. Measures of dispersion
         3. Shape of the data’s distribution
         4. Is there a large amount of missing data? If so, how does this impact your analysis? Explain.

**Part Two: Stakeholder Requirements and Descriptive Analysis Model Application**Using the cleaned and transformed dataset from Part One, you will run the data through descriptive analysis models in Tableau and suggest ways to address the stated business problems.

1. **Using the Stakeholder Requirements Document, identify stakeholder requirements.**
   1. How will these requirements shape your analysis? Explain.
2. **Compare various descriptive analysis methods.**
   1. Compare and contrast a k-means cluster analysis method to at least two of the following models:
      1. Regression
      2. Decision tree
      3. Random forest
   2. Which model is best suited for your analysis?
      1. How does the output of the model support your analysis? Explain.
      2. Why are the other models not suited for your analysis? Explain.
3. **Conduct a k-means cluster analysis.**
   1. Conduct at least three different cluster models and explain the differences.
   2. Explain which model best fits the data. Justify your reasoning.
      1. Which cluster created the most discrete groupings? Explain.
      2. How do these groupings inform your analysis?
4. **Based on your analysis, produce summary statistics for each cluster**.
   1. Explain the subgroups that are present.
   2. Compare and contrast the various subgroups.
   3. Are there any areas where the clusters overlap one another?
   4. How well does your analysis support or meet stakeholder needs?

**Part Three: Analysis and Recommendations Presentation**Use the cleaned and transformed data set to present visual solutions to the various stakeholders. Include any previously created visualizations to support the narrative.

1. **Current State Analysis (Slides 1–4)**
   1. Address the following for both XYZ and ABC restaurant chains, as well as the merged company.
      1. What are its most/least popular items?
      2. What are the most and least profitable items?
      3. Who are the top and bottom customers, in terms of profit?
2. **Findings From Analysis (Slides 5–8)**
   1. Demographic and purchasing profiles of the customer segments
      1. How do the customer demographics compare to the population demographics of the area in which the restaurant chains are located? Because of Incomplete data we made our comparison based on customers within a region who visited the restaurants. The highest visits were to the South with a total of 11k customers and the least visited region was the Northeast with 0.89k visits.
3. **Implications and Recommendations (Slides 9–15)**
   1. Explain the status of the stakeholder requirements.
      1. Did you have the data you needed? No. there were missing Key information like region, Season, Total Population and even an empty item cost for XYZ company. Explain how that impacted your analysis. This has affected the accuracy of my results which might lead to misleading or inaccurate result. I also found it difficult to draw meaningful conclusions leading to high risks of making poor biz decisions.
      2. Explain the relevance and potential limitations of the data underlying your visualizations. Good Data has the following characteristics: 1. Current.2 Complete.3. Consistent to be able to provide valuable insights for any analysis. Possible potential biases /gaps in the data set actually skewed my results limiting the reliability of my visuals and also the conclusions drawn from it.
   2. Recommend areas for improvement for various stakeholders based on your analysis.
      1. Explain shortcomings in the data that may impact your recommendation. There are inconsistency in the data format like the date, time and currency which lead to errors. Also, some records were missing from Company xyz sales record. I also found duplicate data entries which inflated the metrics.
      2. Are there business processes that need to be altered to obtain the correct data? Yes. I believe a uniform format for data entry should be implemented to reduce inconsistencies. The IT unit should introduce automated checks to catch missing or incorrect data and a regular audit should be carried out on data quality to identify and rectify issues.
      3. How does your analysis impact optimal resource allocation for each stakeholder group in the next three years?
         1. Sales: This department had an incomplete sales data and this hampered accurate performance tracking and I would advise that the sales department improves sales tracking system and standardized reporting. With regards to resource allocation, they should focus more on training their sales staff to use new systems effectively and allocate a budget for advanced sales tracking tools to enhance data accuracy. This would lead to better performance and goal setting.
         2. Marketing: I would suggest that the Marketing team would invest more on their customer relationship management system to better understand and target customers. Also allocate resources for training marketing teams on new tools and data usage
         3. Information technology. Advice… the IT department is one of the most important departments in this organization for this reason I advise that you allocate a budget for cyber security measures. Focus on hiring skilled IT personnel to manage and maintain the new systems. This would lead to a unified data platform, improve data accessibility and security.
         4. Finance: All financial transactions should be thoroughly cross checked. Also you should allocate funds for regular financial audits, staff trainings and a good financial software to improve data accuracy

What to Submit

To complete this project, you must submit the following:

**Part One: Transforming Data and Identifying Trends**This should be a 3- to 5-page report, including Power BI screenshots. This should be a Word document with double spacing, 12-point Times New Roman font, and one-inch margins. Sources should be cited according to APA style.

**Part Two: Stakeholder Requirements and Descriptive Analysis Model Application**This should be a 3- to 5-page report, including Tableau screenshots. This should be a Word document with double spacing, 12-point Times New Roman font, and one-inch margins. Sources should be cited according to APA style.

**Part Three: Analysis and Recommendations Presentation**This should be a 15-slide PowerPoint presentation, including Tableau and Power BI screenshots. Sources should be cited according to APA style.

**Challenges with poor data set:**

1. **Incomplete Data: Missing key information hinders accurate analysis.**
2. **Inconsistent Data: When there are variations in the data set, it affects reliability**

**IMPACT ON THE ANALYSIS**

1. **The results generated might end up misleading and or incorrect.**
2. **It was difficulty drawing meaningful conclusions.**
3. **The risk of making poor business decision based on flawed data is high.**

**RELEVANCE AND LIMITATIONS OF DATA**

**RELEVANCE: Data must be current, complete and consistent provide valuable insights.**

**LIMITATIONS: Potential biases and gaps may skew results, limiting the reliability of the visualization and conclusion drawn.**