Analyst Boot Camp Capstone Project 2023

Congratulations, it is now time to consolidate all that you have learned throughout the Analyst Boot Camp into a final capstone project. Parts 1 through 4 will walk you through the steps of your analysis, with a final Part 5 presentation given as a group to stakeholders of the boot camp.

You will be broken into groups to collaborate on the following. All participants are expected to contribute and will be evaluated on each of the project's deliverables.

1. Each of you will be given a different question to answer using BISS data. Determine the scenario to research. Yes, different groups can choose the same scenario. Initial datasets can be found in SQLServer.SQLTraining.BISSBIDSSqlTraining tables in Rick BI
2. Conduct initial research to brainstorm your approach for the business question.
3. Generate hypotheses.
4. Dive into the data and develop your code to answer the question AND make a recommendation.

Introduction

This project encompasses everything covered from the following topics:

1. Data Literacy
2. Relational Data/SQL
3. Data Visualizations
4. Excel & the Analysis Toolpak
5. Algebra & Statistics
6. Business Acumen
7. Data Structure, Parameters, and Stability Indicators

Scenarios

1. What are the top 5 states with the highest average active customers by product?

Based on the data what is your recommendation for states not in the top 5?

1. Rank stores by performance by average sales and product category.

Based on the data what is your recommendation for stores performing poorly?

1. Who are the top 3 agents with the highest average quotes, average customers, and average policies, by state?

Based on the data what is your recommendation for agents performing poorly?

1. What region has the poorest performing staff in terms of total sales and average customers?

Based on the data what is your recommendation for these regions to improve?

1. Rank managers by store performance by number of staff and average sales per customer.

Based on the data what is your recommendation for these regions to improve?

1. Who are the top 5 agents in each state by total policies and close rate (sales/quotes)? What is the top region?

Based on the data what is your recommendation for poor performing agents to improve?

1. What are the 3 bestselling products by total sales by store by region?

Based on the data what is your recommendation for poor performing stores to improve?

1. The recommendation will require you to compare two groups to determine top performers behaviors from poor performers behaviors.
2. The recommendation may require you to answer questions your customer did not ask directly.

Overarching Questions

1. What trends did you find in the data?
2. What alternate joins could be made to the data to answer a different hypothesis?

Building Data Sets

You may have your own hypotheses on your given scenario and dataset, but it may require additional data to support.

Part 1: Create Your Project Management Plan

Using your PMP software of choice (Jira, Trello, Excel spreadsheet, etc.), create a project management plan that shows how you will approach this project. Ensure that your problem statement has the appropriate level of complexity to build a solid presentation.

Once you have created your PMP, meet with your training assistant to approve your initial plan.

Part 2: ETL Documentation

Create a document that explains all ETL involved in your project

* Your document should include numbered steps whenever the order of operation matters.
* Your ETL should be able to be repeated by another person.

Part 3: Napkin Drawings

1. Create [napkin drawings](https://www.youtube.com/watch?v=kuA_yz7aTo0) of visualizations for your final report.
2. Have at least one peer in your class review your visualizations and provide you with feedback (you will need to include this feedback in your final report, so make sure it is documented).
3. Incorporate that feedback into your final visualization.

Part 4: Visualizations

1. Create a dashboard in Excel (or your chosen software) based on your napkin drawing.
2. Your dashboard should have one tile that introduces the data set and links to the data.
3. You should have at least one visualization per data set, with all axes and charts labeled clearly.
4. You should have at least one visual representation of how you filtered your data.

Part 5: Report

Create a project report that includes the following:

* Identify the questions being answered, including the starting questions in the scenario.
* Introduce the datasets and cite them accordingly (ensure that you can confidently explain your dataset).
* Describe what was discovered in the research.
* Support your descriptions with visualizations.
* Explain what your visualizations represent and why they are relevant to the proposed questions.
* Include a correlation matrix or another table showing the Analysis Toolpak from Excel.
* Address any decisions that can be made from this data.

Deliverables

* Presentation of process, including:
* Project Management Plan
* ETL documentation
* Napkin drawings of visualizations
* Document of napkin drawing of visualizations feedback
* Dashboard created in Excel or PowerPoint
* Project Report, Recommendations, & Results

Evaluation Rubric

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Criteria** | **Non-Performance (0)** | | **Basic (1)** | | **Proficient (2)** | | **Distinguished (3)** |
| **Presentation** |  | |  | |  | |  |
| Situation/Task Explanation | Problem is not explained. | | Problem is explained without impact to business identified. | | Impact to business is identified, but problem is not explained. | | Problem is explained with impact to business identified. |
| Presentation Materials | Materials are not proofread or organized. | | Materials have several grammatical errors and poorly organized. | | Materials have minimal grammatical errors and fairly organized. | | Materials are free of grammatical errors and logically organized. |
| **Concepts** |  | |  | |  | |  |
| Analysis Efforts | Only raw data is presented. | | Analysis is presented without data development. | | Presentation includes partial raw data. | | Data are developed and analysis is presented. |
| Statistics Application | Statistics and mathematics not applied to analysis | | General mathematics calculated. | | General mathematics applied to analysis. | | Analysis results are based off statistical application. |
| Data Quality | | Data does not actively meet fit for purpose. | | Data selected is a mix of fit for purpose and unrelated data. | | Data is selected based on fit for purpose. | |
| Impact of Analysis | Recommendations are not made. | | Recommendations are made, but do not include impact of analysis. | | Impact of analysis included, but no recommendations made. | | Strong recommendations based on analysis results and impact. |
| Data Visualization | Results are cluttered and not visually appealing. | | Results are cluttered and minimally visually appealing. | | Results are clear, but not visually appealing. | | Results are clear and visually appealing. |
| Storytelling Analysis Results | Analysis results are not communicated. | | Analysis results are not explained logically. | | Analysis results are minimally explained. | | Analysis results are explained logically. |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Individual** |  |  | |  | |  |
| Presentation Skills | Not all team members participate in presentation. |  | | | | All team members participate in presentation. |
|  |  |  |
|  |  |  | |  | |  |
| Task Completion | | | **Complete (Y/N)** | |  |  |
| Project Management Plan | | |  | |  |  |
| ETL Documentation | | |  | |  |  |
| Napkin Drawing | | |  | |  |  |
| Visualizations | | |  | |  |  |
| Report/Presentation | | |  | |  |  |