Assignment:

Data Warehousing and Profitability Analysis

This is a group assignment. One submission per groups is sufficient.

Part 1)

The vending machine company that you worked on in class is planning to perform a profitability analysis. They have a data warehouse with three dimensions and a fact table. The dimensions include "Time" (daily transactions for two months), "Machine" (vending machines and their locations), and "Product" (drinks and snacks that go into the machines). The fact table includes information about the stales removed from each machine and the new products added to the machines.

The company provides you with a dataset to see how you as a business analyst perform profitability analysis. The dataset is in Excel format, not a real data warehouse format, to make it easier for you to complete the task. The data period involves two months from Dec 2018 to Jan 2019. These are new vending machines, first loaded on 2nd Dec 2018. Also, according to the company, only snacks expire within weeks. Now, they want you to use Excel and its functionalities including pivot tables and charts to complete the following tasks and answer the corresponding questions:

1. Perform a profit analysis by machine for the machines at WPI. Which machines/locations are most profitable? Which machines/locations are least profitable? Explain. (hint: You will need to use an appropriate formula to calculate profit in Excel.)
2. How does usage (represented by stocking and stales) differ for machines at malls vs. companies vs. schools (WPI)? Explain.
3. Do machines differ by the number of stales? Which ones have the higher number of stales? Which ones have lower number of stales? Explain.
4. What are the managerial implications of these results? In other words, based on the results of your analysis, what are your three major recommendations to the manager of the vending machine company to improve their profit? Explain.

Part 2)

Based on the case study titled "Data Warehousing and Multi-Dimensional Data Modelling", which is available in the HBSP course pack, answer the following questions:

1. What is the need for data warehousing at Acme? Explain clearly in a paragraph.
2. Design a multi-dimensional model (i.e., star/snowflake schema) for the *retail product returns* process (see the data warehouse business matrix in that article). Each table/entity, including the dimensions and the fact table, should have at least three attributes including a primary key (and a foreign key, if needed). You can simply use *Shapes* in Word to draw the model. However, if you decide to use any other tool, just provide an image or screenshot of the model here. For the connections (relationships) between entities in the model you can use simple lines; no need to add other details to the connections or lines.

Part 3)

Based on the ERD that we used in class for the queries, write two SQL statements/queries that answer the following question: Which five products were the most successful in the month of January (2020) in terms of:

- *Number of items sold* [Return product name/description along with the number of items sold for each of the top five products]
- *Total sales ($)* [Return product name/description along with the total sales ($) for each of the top five products]

What to submit: You will submit one file:

1. A Word or PDF document that contains answers to the questions in each part.