Report_Draft_Summary_v5

December 5, 2024

1 Summary

- ?? high-level summary of conclusions
- ?? report objectives etc??

2 Objectives & Approach

This report describes the findings from an analysis of sales review data for the major US retail chain, JCPenney. It is the final assignment for the module ITNPBD2 - Representing and Manipulating Data. This analysis uses the five datasets provided, with some supplementary data to augment.

The report is structured into four sections:

- Summary
- Data Collation, Exploration & Preparation
- Data Analysis
- Conclusions

2.1 Approach

This is intended primarily as a business report to identify and make recommendations for improvements; with appropriate supporting evidence. In addition, there is a need to meet the additional assignment objectives of:

- Demonstrate that a data science process (here CRISP-DM) has been followed
- Demonstrate the use of Python for data manipulation and analysis

However, once the analysis commenced it quickly became evident that two challenges needed to be addressed. The first is that trying to produce a single PDF report with commentary, Python code and runtime output would result in a very long and difficult to read artefact. The second challenge was that a brief examination of the data using Data Wrangler (Microsoft, 2023) quickly revealed that the data was of very poor quality and would make any meaningful business conclusions possibly meaningless.

As a result, the following approach has been taken:

- Focus on the data preparation part of the process and make recommendations on areas to improve
- Carry out some limited business analysis, but really to demonstrate techniques and not make any business recommendations

- From a process and coding perspective, establish a set of processes that can be reused when better quality source data can be provided

And in terms of assignment documents:

- Submit a report PDF with limited Python code
- Submit the set of Jupyter Notebooks used to produce the above

2.2 CRISP-DM

The approach to the analysis follows the first four parts of the six stage "CRoss-Industry Standard Process for Data Mining' (CRISP-DM) process (Ncr et al., 1999), (Hotz, 2024). In summary, this process is:

1. Business Understanding: Define project objectives and requirements by collaborating with stakeholders
2. Data Understanding: Collect and explore data, analysing its characteristics and quality
3. Data Preparation: Clean, handle missing values, and transform variables to create a structured dataset
4. Modelling: Apply various techniques such as machine learning algorithms or statistical models to the prepared data
5. Evaluation: Rigorously assess models based on predefined criteria, including performance and reliability
6. Deployment: Integrate successful models into existing systems and monitor their effectiveness

2.3 Who Are JCPenney?

JCPenney (JCP) is a major North American department store chain Wikipedia, operating as Penney OpCo LLC. JCP has 656 stores in 49 states plus Puerto Rico according to the JCP Store Locator.

In 2020, JCP filed for bankruptcy and were purchased by an asset management company, a large number of stores were also closed. Later, in August 2023, JCP announced a major turnaround plan to replace its current website and inventory management systems, as well as make major upgrades to its retails stores (ModernRetail, 2023).

3 Conclusions

4 References

Microsoft (2023) https://devblogs.microsoft.com/python/data-wrangler-release/

Ncr, P.C. et al. (1999) 'CRISP-DM 1.0'

Hotz (2024) https://www.datascience-pm.com/crisp-dm-2/

Modern Retail (2023) https://www.modernretail.co/operations/jcpenney-is-the-latest-department-store-to-announce-a-major-turnaround-plan/

xxxx

Spiegel, M.R. and Schiller, J. (2012) Schaum's Outline Probability and Statistics. 4th edn. McGraw Hill.