678 proposal

yingmai chen

November 8th 2023

1 Personal Statement

My career goal is to be a data scientist, I am particularly drawn to the intricate dance of market trends, where numbers and narratives converge to reveal the pulse of commerce. It is here, in the meticulous dissection of market fluctuations and consumer behaviors, that I seek to establish my expertise. My ambition is to harness the predictive power of data analytics to forecast and shape market trajectories. This mid-term project, which centers on analyzing Walmart's sales data from 2010 to 2012, is a deliberate step toward this goal.

2 Research objective

The objective of this research is to apply multilevel modeling[2] techniques to understand how various store-level characteristics and regional economic indicators influence weekly sales across Walmart stores.And the ultimate goal of study is to identify key factors that drive sales performance at different levels – store and regional – to aid in strategic decision-making for resource allocation and marketing strategies. And the ideas come from[1].

3 datasets

The data comes from Kaggle (https://www.kaggle.com/datasets/yasserh/walmart-dataset),contains 6,435 entries and 8 columns, which include:

Store: An identifier for each Walmart store.
Date: The week for which sales data is provided.
Weekly Sales: The total sales for the given store in the given week.
Holiday Flag: A binary indicator of whether the week includes a holiday (1) or not (0).
Temperature: The average temperature in the region of the store in Fahrenheit.
Fuel Price: The cost of fuel in the region.
CPI: The Consumer Price Index for the region.
Unemployment: The unemployment rate for the region.

4 Timeline

EDA: November 9th to November 13th
Data processing: November 14th to November 20th
Modeling and Validation: November 21st to November 26th
Write up: November 27th to December 7th

References

[1] Zhaoyu Chen. Sales forecast of walmart on account of multivariate regression and machine learning methods. In *Proceedings of the International Conference on Financial Innovation, FinTech and Information Technology, FFIT 2022, October 28-30, 2022, Shenzhen, China*, 2023.

[2] Joop Hox. Multilevel modeling: When and why. In *Classification, data analysis, and data highways: proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation eV, University of Potsdam, March 12–14, 1997*, pages 147–154. Springer, 1998.