## **Machine learning and predective analysis with movie data**

Author: Jinhang Jiang

### Introduction

This machine learning project is the final project for the Business Analytics Capstone Class (BSAN 480, INSTR: Chris Claterbos) at the University of Kansas School of Business. The purpose of the project is to study patterns and derive insights from the ML Movie data and IMDB data to predict the performance and the rating of a movie in the future.

### Scope of Work

For the main concentration, the project will focus on performance analysis and prediction. I firstly cleaned, created, and then analyzed the given tables (Appendix A-1 & A-2) to understand patterns and stories that otherwise may have gone untold. Secondly, I created a tag table (Appendix A-3) and did a sentimental analysis with the data. I also created some dashboards to visualize the significant findings. Finally, I tried and compared several different predictive models to forecast profitability performance, as well as the ratings.

I used SQL, R, and Python as the programming languages, Node2Vec, Networkx, and K-means for analysis, Rstudio, Jupyter Notebook, Oracle Developer as IDE, and Oracle Analytics Cloud to create dashboards.

### Data

Identify the data, where it came from. Include the data model, entities and key attributes, Identify base measures and key calculated measures. If the data has joins explain the join conditions. (use figures where necessary)

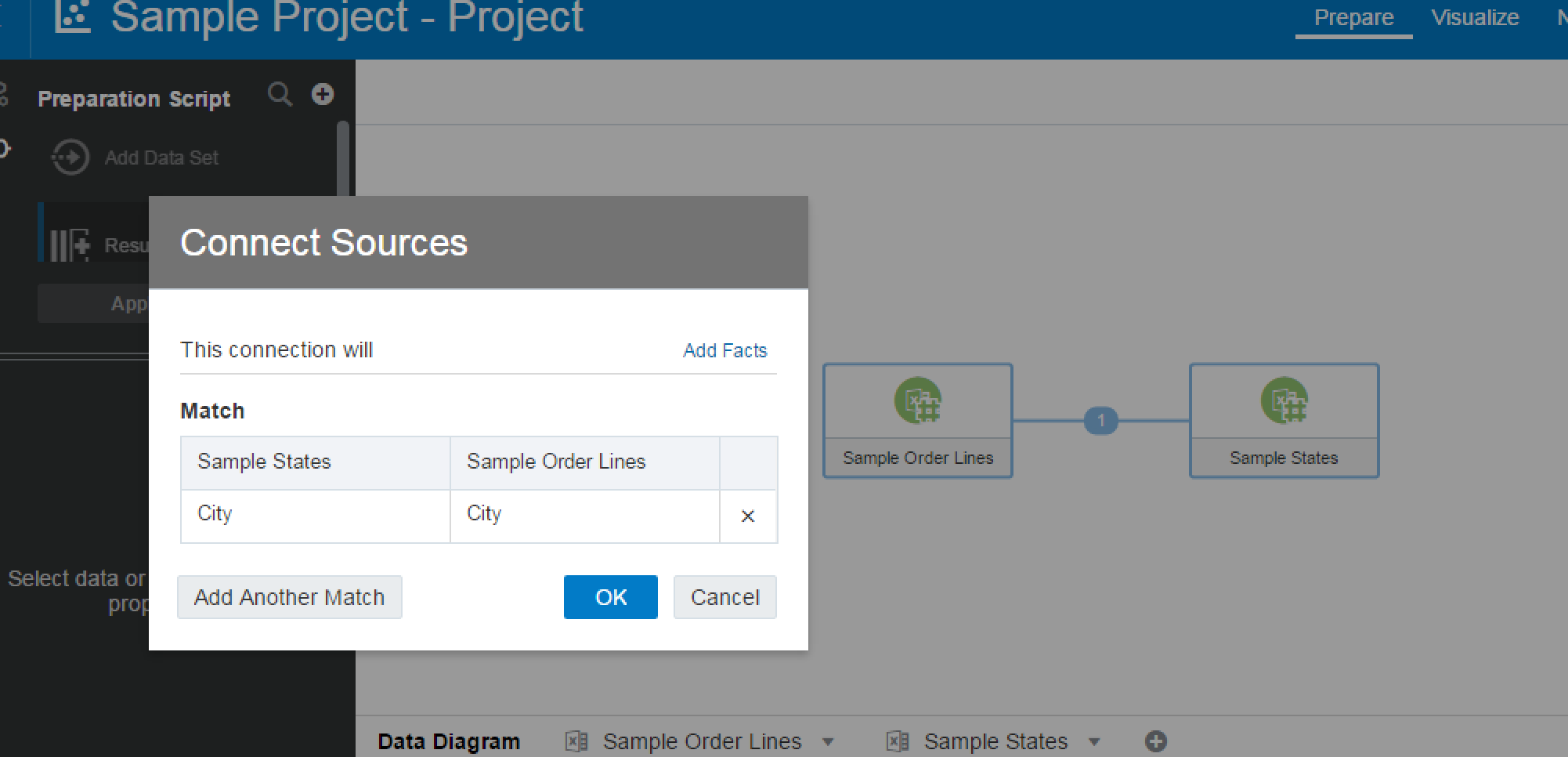


Figure - Data Model A

### Analysis Process

What was done….Preliminary analysis

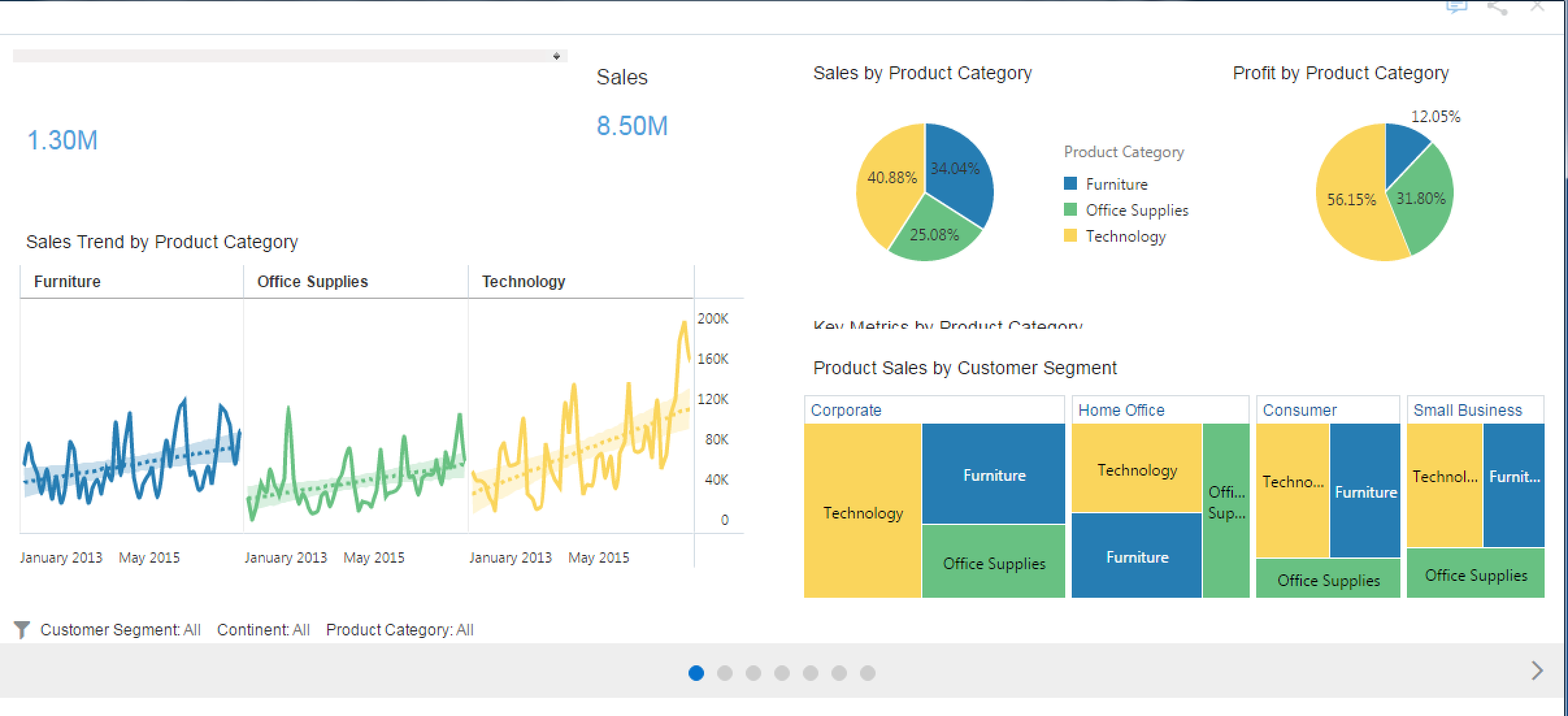


Figure - sample dashboard

### Results

What insights and results were found.

### Conclusion

Summary of work and insights / recommendations.

# APPENDIX A

You may check the partial codes and relevant tables on my [GitHub](https://github.com/jinhangjiang/BSAN480-FinalProject).

1. [SQL code](https://github.com/jinhangjiang/BSAN480-FinalProject/blob/master/Scripts/MovieTable_sqlcommand.txt) for cleaning and creating the original table.
2. [R code](https://github.com/jinhangjiang/BSAN480-FinalProject/blob/master/Scripts/Model_analysis.R) for modifying the movie table.
3. [SQL code](https://github.com/jinhangjiang/BSAN480-FinalProject/blob/master/Scripts/Tag_creation.sql) for creating the tag table.
4. Dictionary and references for the tables