

Progress Report

Due: 10/24 10:00 AM (EST) 星期六

- Write a one to two page report in R Markdown. Include the following elements:
 - App Title and Names of your group members
 - Description of the problem to be analyzed in the Shiny App (1) - Xubo & Jiarong
 - Proposed data sources. (1) - Xubo & Jiarong
 - Concept for the Shiny App
- * Overall Layout (1) - Yunting
- * Data Analysis and Visualization with User Choices and Results (2) - Yunting
- * Statistical Modeling with User Choices and Results (2) - use a prices.offer variable to compare online and in-store price - Shan Lin
- Allocation of Responsibilities for the team (1) - Shan Lin
- Project Steps/Schedule (1) - Shan Lin

10.27	Project report
11.3	Tidy data
11.10	Prepare our questions and Analyze data
11.17	Deepen the analysis
11.24	Rejust data
12.1	Prepare our presentation
12.8	upload analysis and give a presentation

#Group Member And Assigned Roles:

- Xubo
 - Tidying Data, Data analysis, Final Presentation
- Yunting Chui
 - Tidying Data, Data analysis, Final Presentation
- Jiarong Li
 - Report writing, Data visualization, Final Presentation
- Shan Lin
 - Report writing, Data visualization, Final Presentation

Shiny App's title

The analysis of Walmart Women's Shoes between March to May 2019.

Group Member:

1. Jiarong Li
2. Shan Lin
3. Xubo Tang
4. Yunting Chiu

Our analysis in Shiny App

With the improvement of women's economic ability and the variability of fashion demands, women's shoes have become an important part of the footwear market as their demands for shoes increased. Some problems with meaningful business insights are brought to the attention of related parties, such as analysts, vendors and their customers: What is the average price of each distinct brand? Which brands have the highest prices? Which one has the widest distribution of prices? Is there a typical price distribution (e.g., normal) across brands or within specific brands? Is there an obvious difference in price between online and in store? The Shiny App of this project aimed to answer these questions and is used to determine brand markups, pricing strategies, and trends for luxury shoes.

Proposed data sources

The dataset is from the Data. World and created by Datafiniti. The dataset was updated 2019 which is a list of 10,000 women's shoes. There are 29 variables which contain price, brand, manufacture, ID, price discount, price size, shoes size, date created, image URL and so on. After going through the dataset, there are some variables that can be analyzed: price, brand, manufacture and price offer.

Concept for the Shiny App

Overall Layout

According to the dataset, we find out some interesting variables. We will show the research results that we could derive from the data following the below questions.

- What is the relationship between brand and manufacturer? Are they the same?
- How do companies price their products? (consider the prices.availability)
- Should retailers match their own Prices online and in stores?
- Which retailer has the most discount?

Data Analysis and Visualization with User Choices and Results

Data Analysis in dataTableOutput:

- Use the prices.offer variable to compare online prices and in-store prices.
- Which brand has the highest price, on average.
- Find out the popular colors of women's shoes with descending order.
- Show the relationship between brand and manufacturer (true or false).

Data Visualization in plotOutput:

- Users can choose the retailers to see how they decide the price to women shoes.
- Choose the important categorical variables to make a barplot function (y axis is count).
- See whether the quantitative variables follow a normal distribution (that is, median is approximately equal to mean).
 1. Before discount prices
 2. After discount prices

·so far;

So far, we have sorted out a large part of the data, but the current data still does not meet our expectations, but there are still a few more confusing dates such as update and dateAdd, and the color classification is messy and not clear enough. It brings inconvenience to our analysis, and the time is only provided from March to May, but we want to divide it by season to analyze sales volume and so on.

·Next step

In the next step, we will continue to beautify the data, visualize the data, and then perform a hypothesis check