Predict Future Sales

Siyu Chen

Xi'an Shiyou University

(None)

Overview

Problem Definition

Data Pre-processing

Data Visualization

Problem Definition

Analyze Data

Data Pre-processing

Deal with two kinds of data

Data Visualization

Problem Definition

Analyze Data

Data Pre-processing

Data Visualization

Problem Definition

Analyze Data

Problem Definition
Analyze Data
Data Pre-processing
Data Visualization

Defn

predict total sales for every product and store in the next month.

- sales_train.csv: Training set (2013 to October 2015)
- test.csv: Test set (Forecast sales of these stores and products in November 2015)
- sample_submission.csv: A properly formatted sample submission file
- items.csv: Additional information about the merchandise/product
- item_categories.csv: Additional information on the categories of goods
- shops.csv: Additional information about the store

File field description

Problem Definition

Analyze Data

Data Pre-processing

Data Visualization

- shop_id : unique identifier for a store.
- item_id : A unique identifier for a product
- item_category_id : A unique identifier for a category
- item_cnt_day : Quantity of products sold

- item_price : The current price of the goods.
- date_block_num : A consecutive month.
- item_name : product name.
- shop_name : shop name.
- item_category_name : The name of the project category.

Problem Definition

Data Pre-processing

Deal with two kinds of data

Data Visualization

Data Pre-processing

Deal with two kinds of data

Problem Definition

Data Pre-processing

Deal with two kinds of data

Data Visualization

Defn

- Erroneous data Change the negative price to the median price of the item in the store.
- Remove individual data Sales of more than 1000 should be excluded.Looking at the price distribution finds that should eliminate values greater than 100000

Figure 1: P

Problem Definition

Data Pre-processing

Data Visualization

Data Visualization

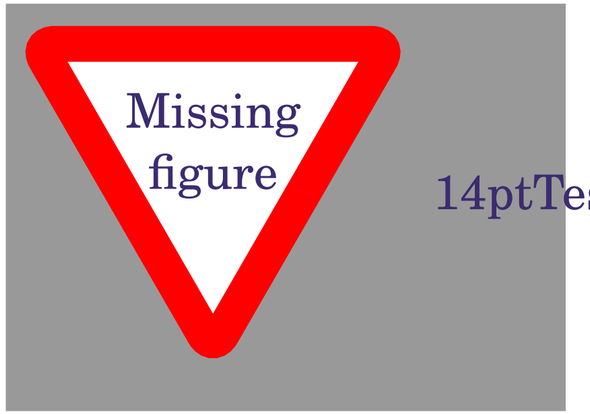
data visualization

Problem Definition

Data Pre-processing

Data Visualization

- To predict next month's sales, find out how other factors affect sales
  - Draw a bar chart according to the category, and the highest selling product is No. 40.
  - Take 12 months as the observed value and draw the trend of monthly mean and sample standard deviation.
  - Plot monthly mean sales and monthly mean standard deviation.



14ptTest

(a) Original Feature Spaces

14ptTest

(b) Group Outlying Spaces

14ptTest.

(c) Another Subspaces

Information acquired

Problem Definition
Data Pre-processing
Data Visualization

- Item 40 sells the most.
- The sales volume in December and 24 months with obvious peak is judged to be the peak in the annual cycle.
- The monthly mean shows a decreasing trend, indicating that commodities have been declining on an annual basis.
- The sample variance increased month by month, indicating that the sales volume declined from the beginning of the second year to the beginning of the second year before stabilizing the downward trend.

Contact Information

Thank you for listening!

Siyu Chen

Xi’an Shiyou University

785987165@qq.com