

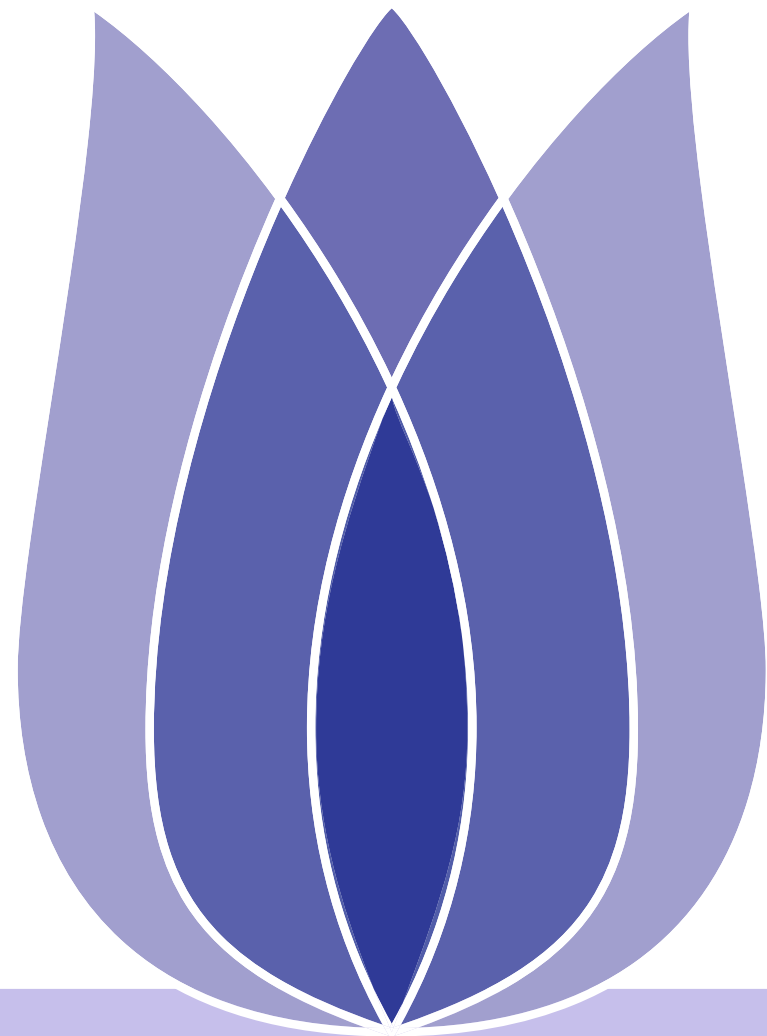


# Predict Future Sales

Siyu Chen

Xi'an Shiyou University

2020-09-28





# Overview



# Problem Definition



**TULIP**

*Team for Universal Learning and Intelligent Processing*



# Analyze Data

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

- **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.





# Data Pre-processing



**TULIP**

*Team for Universal Learning and Intelligent Processing*



## Deal with two kinds of data

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



**TULIP**

*Team for Universal Learning and Intelligent Processing*





# Data Visualization



**TULIP**

*Team for Universal Learning and Intelligent Processing*



## data visualization

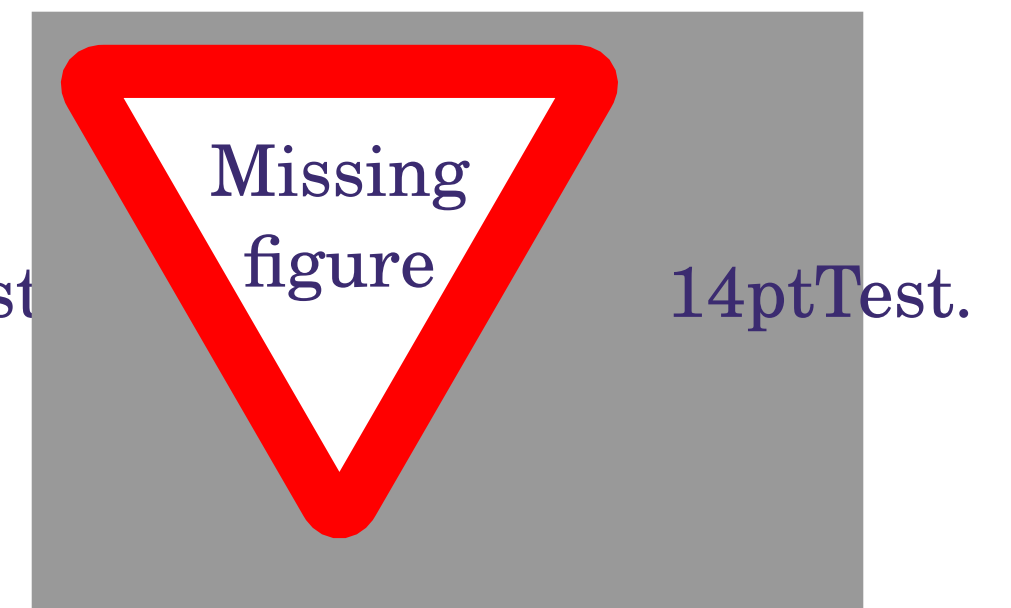
- To predict next month's sales, find out how other factors affect sales
  - ◆ Draw a bar chart according to the category, and the highest selling product is No. 40.
  - ◆ Take 12 months as the observed value and draw the trend of monthly mean and sample standard deviation.
  - ◆ Plot monthly mean sales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



(a) Original Feature Spaces



(b) Group Outlying Spaces



(c) Another Subspaces





## Information acquired

- 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

