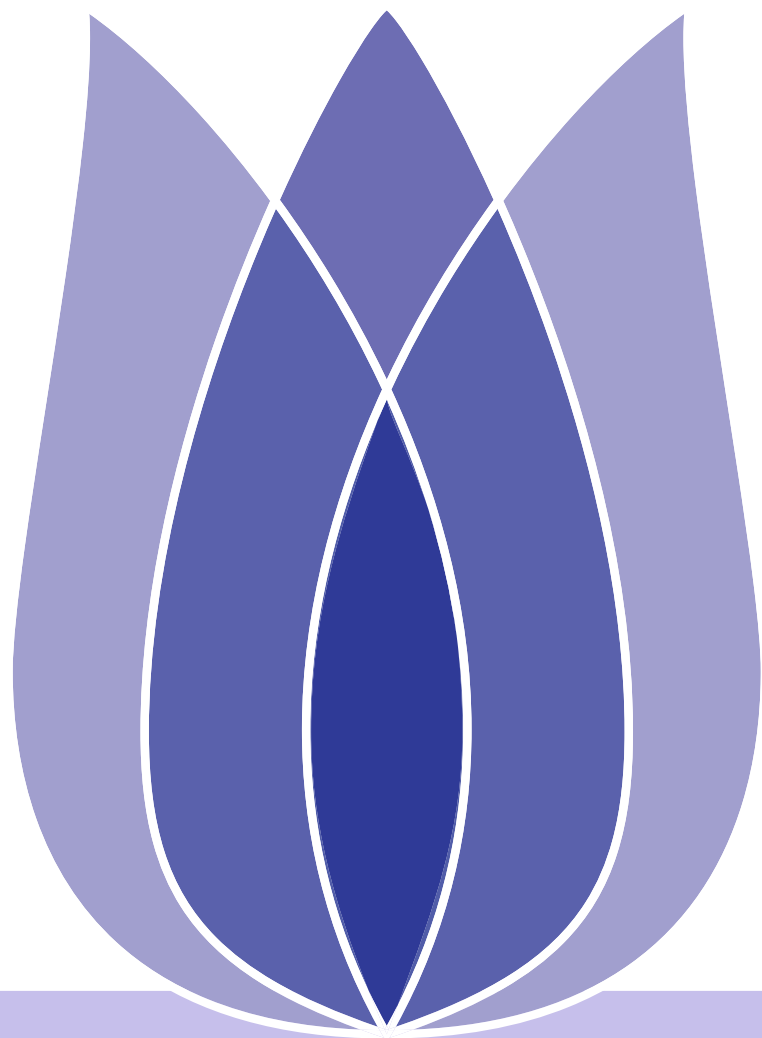




# FLIP(00) Mid-term Presentation

Rongxin Xu  
Hunan University

26 October 2019





# Outline

- [Introduction](#)
- [Data Description](#)
- [Exploratory Data Analysis](#)
- [Stationarity](#)
- [Conclusion](#)

- Introduction**
- Data Description**
- Exploratory Data Analysis**
- Stationarity**
- Conclusion**



- Introduction
- Problem Description
- Data Description
- Exploratory Data Analysis
- Stationarity
- Conclusion

# Introduction



# Problem Description

- Introduction
- Problem Description
- Data Description
- Exploratory Data Analysis
- Stationarity
- Conclusion

- This is a problem with time-series prediction. There are six data sets with a total of 11 attributes. The following is some requirements.



# Problem Description

- Introduction
- Problem Description
- Data Description
- Exploratory Data Analysis
- Stationarity
- Conclusion

- This is a problem with time-series prediction. There are six data sets with a total of 11 attributes. The following is some requirements.
  - ◆ According to the given train data set, training a model and then using the model to predict total sales for every product and store in the next month.



- [Introduction](#)
- [Data Description](#)
- [Attribute Information](#)
- [Detailed description](#)
- [Summary](#)
- [Exploratory Data Analysis](#)
- [Stationarity](#)
- [Conclusion](#)

# Data Description



- Introduction
- Data Description
- Attribute Information
- Detailed description
- Summary
- Exploratory Data Analysis
- Stationarity
- Conclusion

## ■ Attribute Information

1. There are six data sets with a total of 11 attributes.

Attribute name	description
ID	An Id that represents a (Shop, Item) tuple within the test set.
shop_id	Unique identifier of a shop.
item_id	Unique identifier of a product.
item_category_id	Unique identifier of item category.
item_cnt_day	Number of products sold. You are predicting a monthly amount of this measure.
item_price	Current price of an item.
date	Date in format dd/mm/yyyy.
date_block_number	A consecutive month number, used for convenience. January 2013 is 0, February 2013 is 1,..., October 2015 is 33.
item_name	Name of item.
item_category_name	Name of item category.
shop_name	Name of shop.

Figure 1: Attributes name and description

2. The detailed description of the data is shown in the following table.





# Detailed description

- Introduction
- Data Description
- Attribute Information
- Detailed description
- Summary
- Exploratory Data Analysis
- Stationarity
- Conclusion

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0
5	10.01.2013	0	25	2564	349.00	1.0

(a) sales\_train.csv

	item_category_name	item_category_id
0	PC - Гарнитуры/Наушники	0
1	Аксессуары - PS2	1
2	Аксессуары - PS3	2
3	Аксессуары - PS4	3
4	Аксессуары - PSP	4
5	Аксессуары - PSVita	5

(b) item\_categories.csv

ID	shop_id	item_id
0	0	5
1	1	5
2	2	5
3	3	5
4	4	5
5	5	5

(c) test.csv

	item_name	item_id	item_category_id
0	! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
1	!ABBY FineReader 12 Professional Edition Full...	1	76
2	***В ЛУЧАХ СЛАВЫ (UNV) D	2	40
3	***ГОЛУБАЯ ВОЛНА (Univ) D	3	40
4	***КОРОБКА (СТЕКЛО) D	4	40
5	***НОВЫЕ АМЕРИКАНСКИЕ ГРАФИТИ (UNI) ...	5	40

(d) items.csv

	shop_name	shop_id
0	!Якутск Орджоникидзе, 56 фран	0
1	!Якутск ТЦ "Центральный" фран	1
2	Адыгея ТЦ "Мега"	2
3	Балашиха ТРК "Октябрь-Киномир"	3
4	Волжский ТЦ "Волга Молл"	4
5	Вологда ТРЦ "Мармелад"	5

(e) shops.csv

Figure 2: Data Description



# Summary

- Introduction
- Data Description
- Attribute Information
- Detailed description
- Summary
- Exploratory Data Analysis
- Stationarity
- Conclusion

- The data is very clean and complete, so we only need to change the data type after importing.
- We also need to reorganize the table structure to make it more readable. An sample is given below.

			date		item_price	item_cnt_day
			min	max	mean	sum
date_block_num	shop_id	item_id				
0	0	32	2013-01-03	2013-01-31	221.0	6.0
		33	2013-01-03	2013-01-28	347.0	3.0
		35	2013-01-31	2013-01-31	247.0	1.0
		43	2013-01-31	2013-01-31	221.0	1.0
		51	2013-01-13	2013-01-31	128.5	2.0
		61	2013-01-10	2013-01-10	195.0	1.0
		75	2013-01-17	2013-01-17	76.0	1.0
		88	2013-01-16	2013-01-16	76.0	1.0
		95	2013-01-06	2013-01-06	193.0	1.0

Figure 3: sample



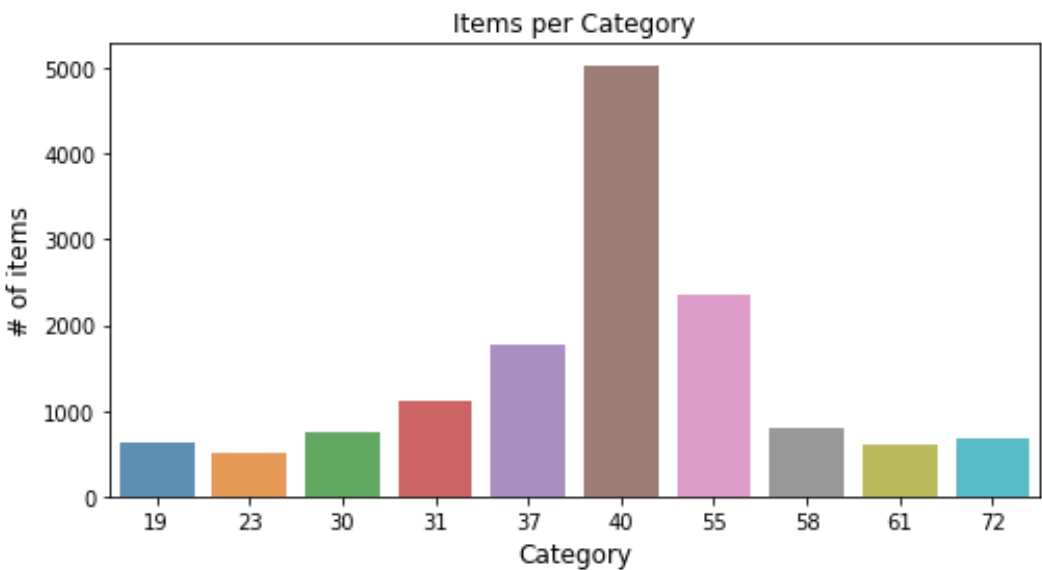
- [Introduction](#)
- [Data Description](#)
- [Exploratory Data Analysis](#)**
- [Exploratory Data Analysis](#)
- [Summary](#)
- [Stationarity](#)
- [Conclusion](#)

# Exploratory Data Analysis

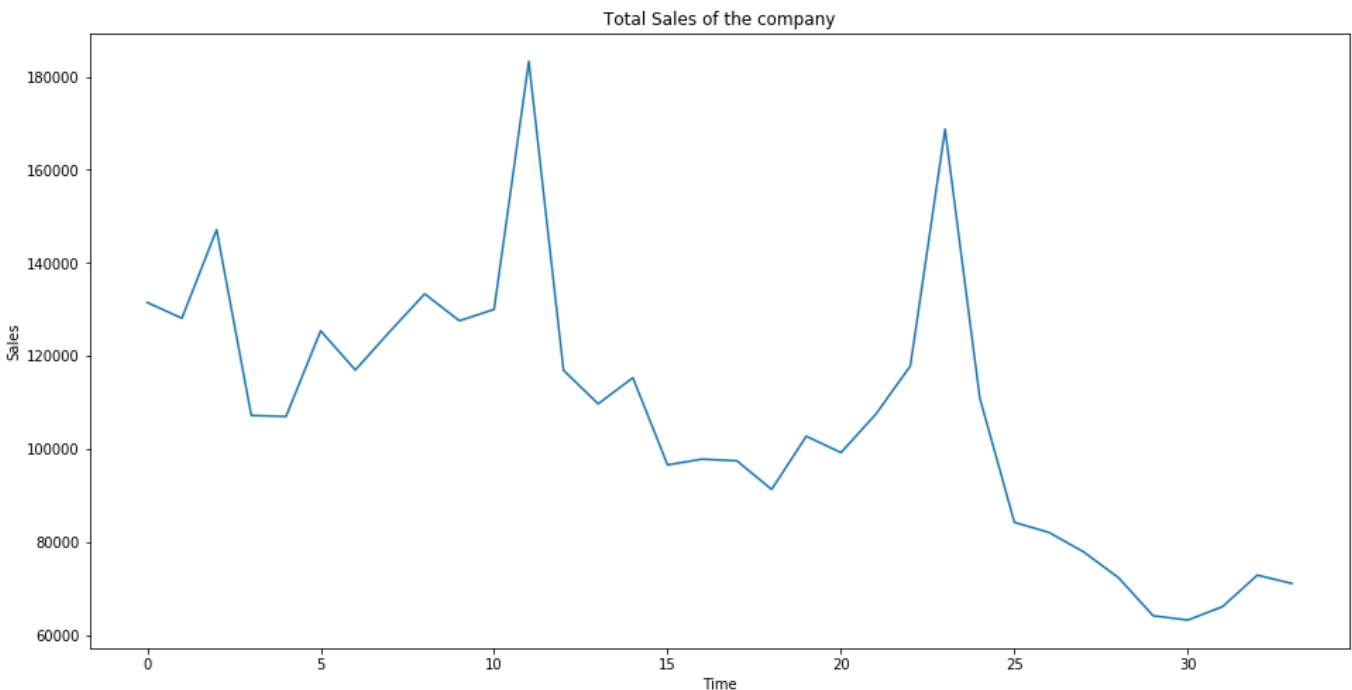


# Exploratory Data Analysis

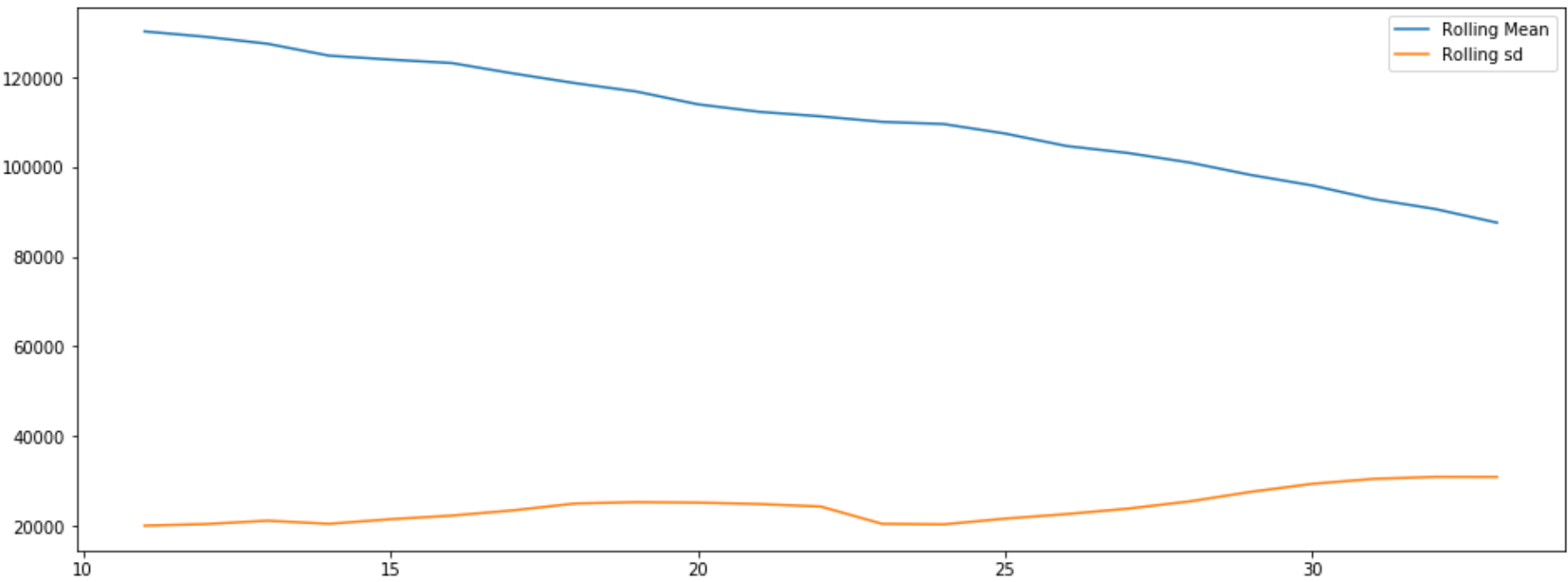
- Introduction
- Data Description
- Exploratory Data Analysis
- Exploratory Data Analysis**
- Summary
- Stationarity
- Conclusion



(a) Items per Category



(b) Total Sales of the company



(c) Rolling Mean and std

Figure 4: EDA



# Summary

- [Introduction](#)
- [Data Description](#)
- [Exploratory Data Analysis](#)
- [Exploratory Data Analysis](#)
- [Summary](#)
- [Stationarity](#)
- [Conclusion](#)

- There is an obvious "seasonality" (Eg: peak sales around a time of year) and a decreasing "Trend".



[Introduction](#)

[Data Description](#)

[Exploratory Data Analysis](#)

**[Stationarity](#)**

[Seasonality and Trend](#)

[Stationarity Test](#)

[Remove seasonality and trends](#)

[Summary](#)

[Conclusion](#)

# Stationarity



# Seasonality and Trend

- Introduction
- Data Description
- Exploratory Data Analysis
- Stationarity
- Seasonality and Trend**
- Stationarity Test
- Remove seasonality and trends
- Summary
- Conclusion

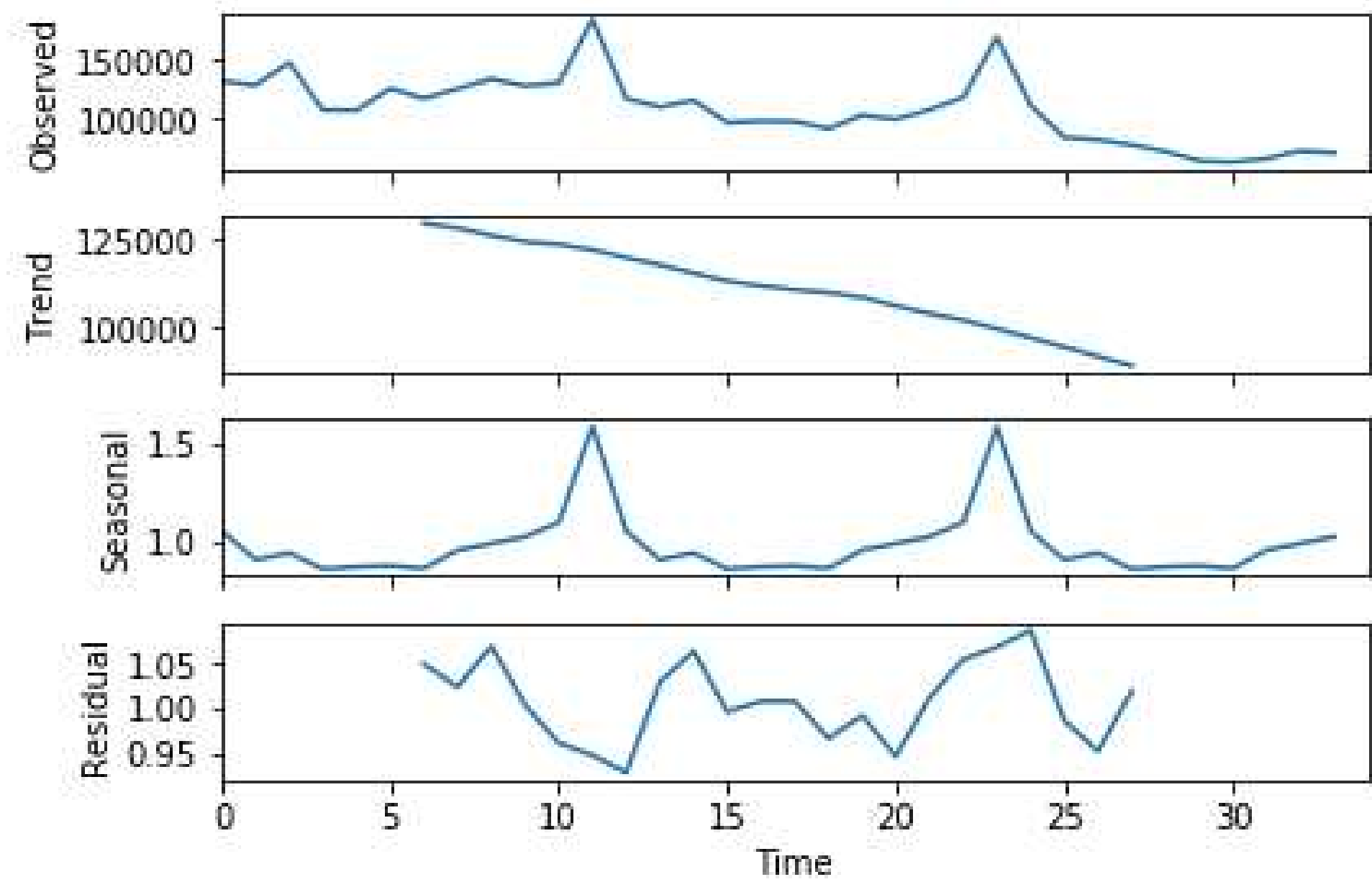


Figure 5: Seasonality and Trend



# Stationarity Test

- Introduction
- Data Description
- Exploratory Data Analysis
- Stationarity
- Seasonality and Trend
- Stationarity Test**
- Remove seasonality and trends
- Summary
- Conclusion

```
Results of Dickey-Fuller Test:
Test Statistic          -2.395704
p-value                 0.142953
#Lags Used              0.000000
Number of Observations Used 33.000000
Critical Value (1%)     -3.646135
Critical Value (5%)     -2.954127
Critical Value (10%)    -2.615968
dtype: float64
```

Figure 6: Stationarity Test





# Remove seasonality and trends

- Introduction
- Data Description
- Exploratory Data Analysis
- Stationarity
- Seasonality and Trend
- Stationarity Test
- Remove seasonality and trends
- Summary
- Conclusion

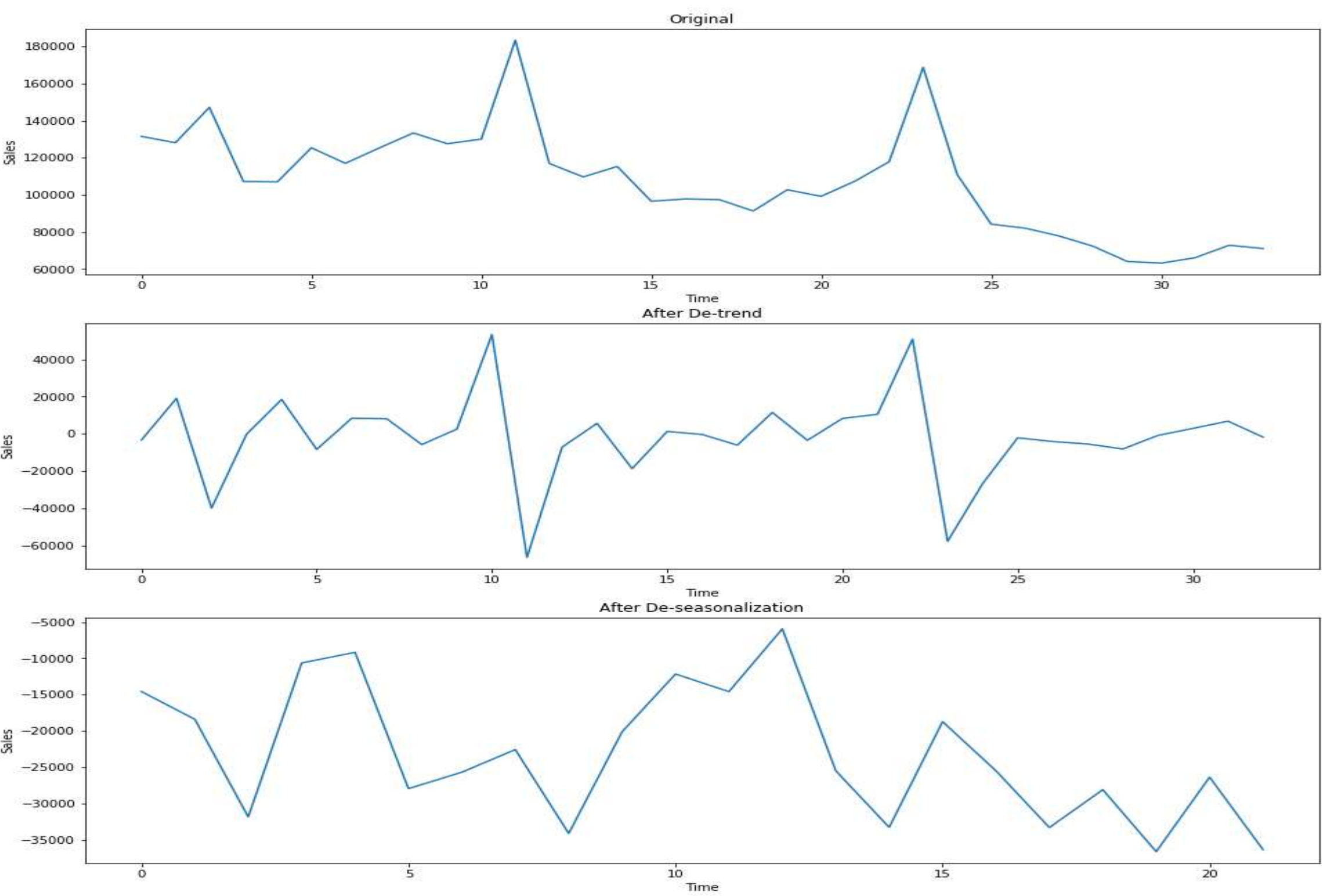


Figure 7: Remove seasonality and trends



- Introduction
- Data Description
- Exploratory Data Analysis
- Stationarity
- Seasonality and Trend
- Stationarity Test
- Remove seasonality and trends
- Summary**
- Conclusion

- Now let’s check the new P-value.

```
Results of Dickey-Fuller Test:
Test Statistic          -3.270101
p-value                  0.016269
#Lags Used              0.000000
Number of Observations Used  21.000000
Critical Value (1%)      -3.788386
Critical Value (5%)      -3.013098
Critical Value (10%)     -2.646397
dtype: float64
```

Figure 8: new stationarity test

- After the transformations, our p-value for the DF test is well within 0.05. Hence we can assume Stationarity of the series.



- [Introduction](#)
- [Data Description](#)
- [Exploratory Data Analysis](#)
- [Stationarity](#)
- [Conclusion](#)**
- [Summary](#)
- [Future research](#)

# Conclusion



# Summary

- [Introduction](#)
- [Data Description](#)
- [Exploratory Data Analysis](#)
- [Stationarity](#)
- [Conclusion](#)
- [Summary](#)
- [Future research](#)

- From the above result presentation, we can find that  
There are seasonality and trend in data.
- From the Stationarity test, we can find that  
After removing seasonality and trends, the time series becomes smooth.  
So we can use traditional time series prediction methods for prediction.



## Future research

Introduction
Data Description
Exploratory Data Analysis
Stationarity
Conclusion
Summary
Future research

- Predict by traditional time series prediction models such as AR, MA and ARMA.
- Using more models to predict, such as random forests and neural networks.
- Find the most effective model and get my own kaggle ranking.





Thank you & Question

