



IGNACIO MALPARTIDA



Sales prediction - Kaggle competition
Business intelligence + Business analytics

Pareto principle

Focus the business effort
on maximizing value,
improving marketing
campaigns, and focusing on
the true growth

20%
CAUSES

80%
EFFECTS

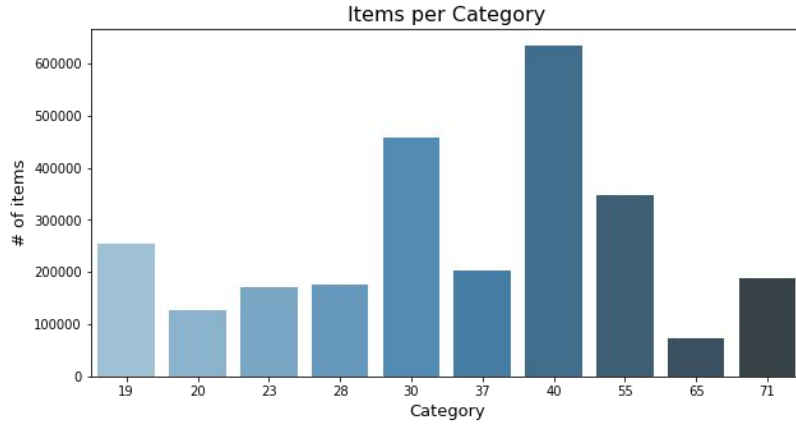


PARETO PRINCIPLE

EDA

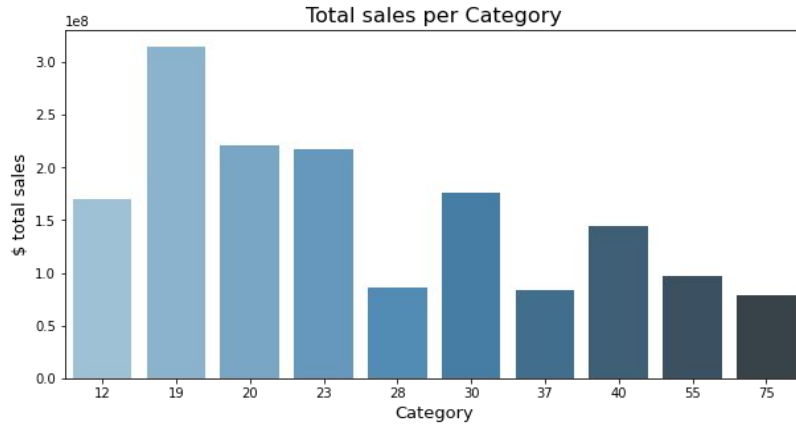
Exploratory Data Analysis

**Which categories are
the most sold?**



The 3 most sold categories are:

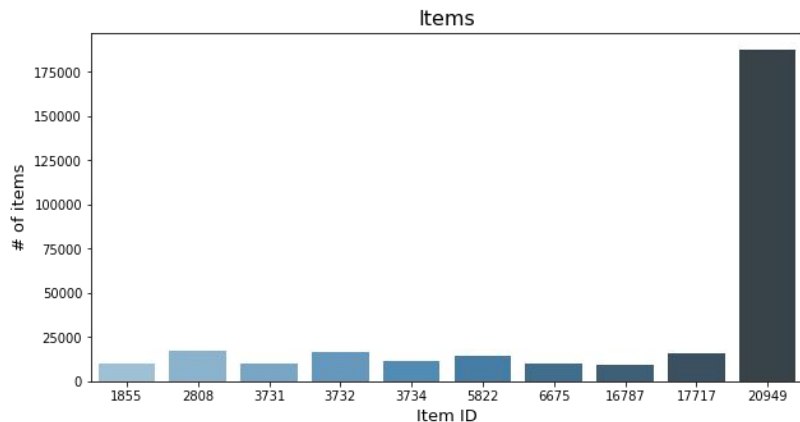
- Category #40
- Category #30
- Category #55



The 3 most valuable \$ categories are:

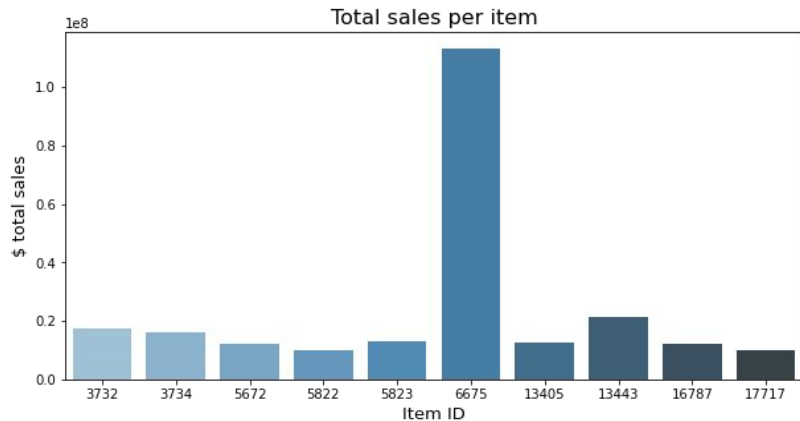
- Category #19
- Category #20
- Category #23

**Which articles are the
most sold?**



The 3 most sold items are:

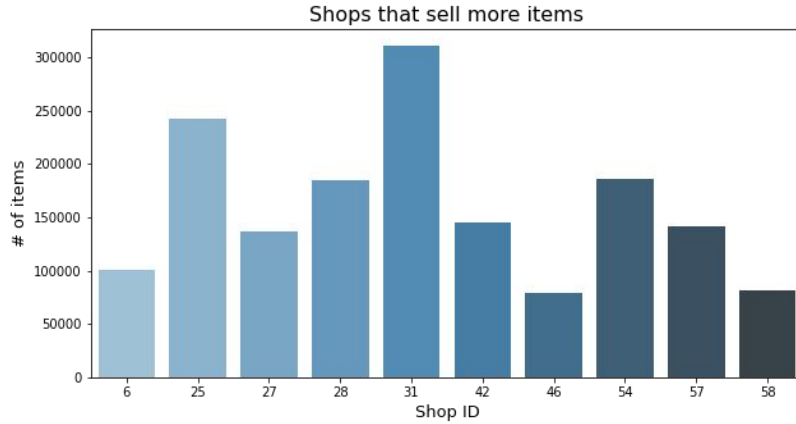
- Item#20949
- Item#2808
- Item#3732



The 3 most valuable \$ categories are:

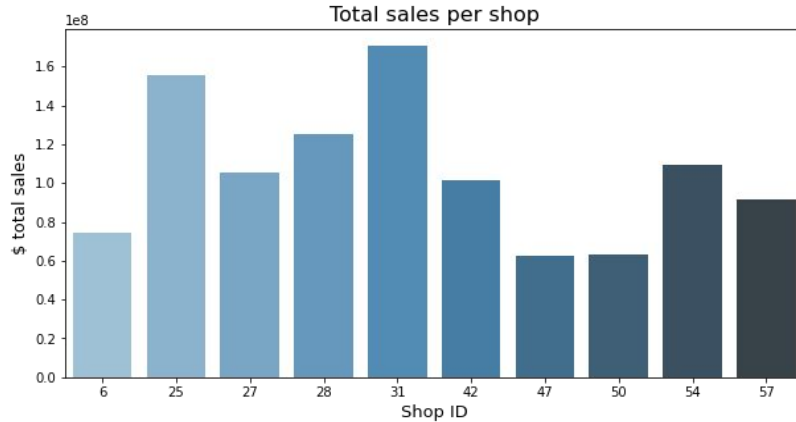
- Item#6675
- Item#1343
- Item#3732

**Which stores sell the
most?**



The 3 most important shops for items volumes are:

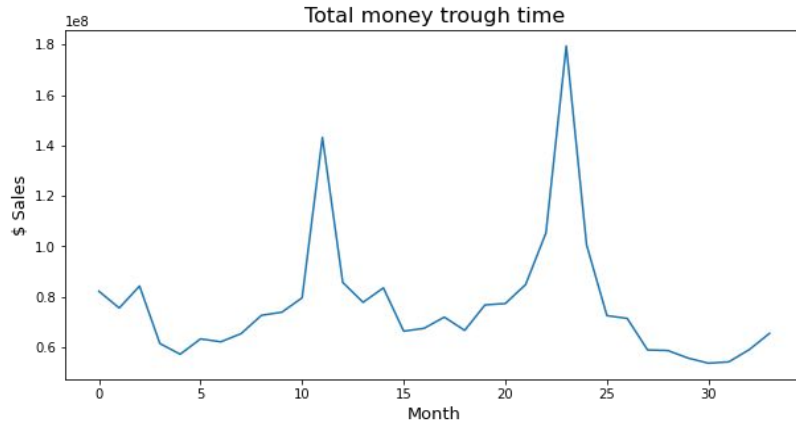
- Shop#31
- Shop#25
- Shop#54



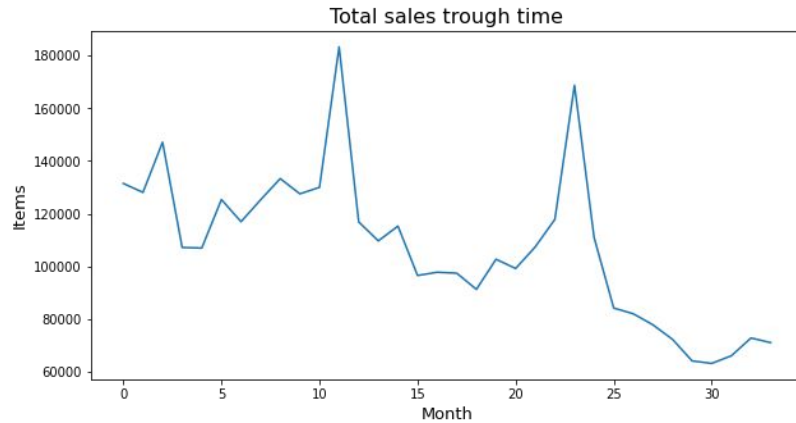
The 3 most important shops for items sales are:

- Shop#31
- Shop#25
- Shop#28

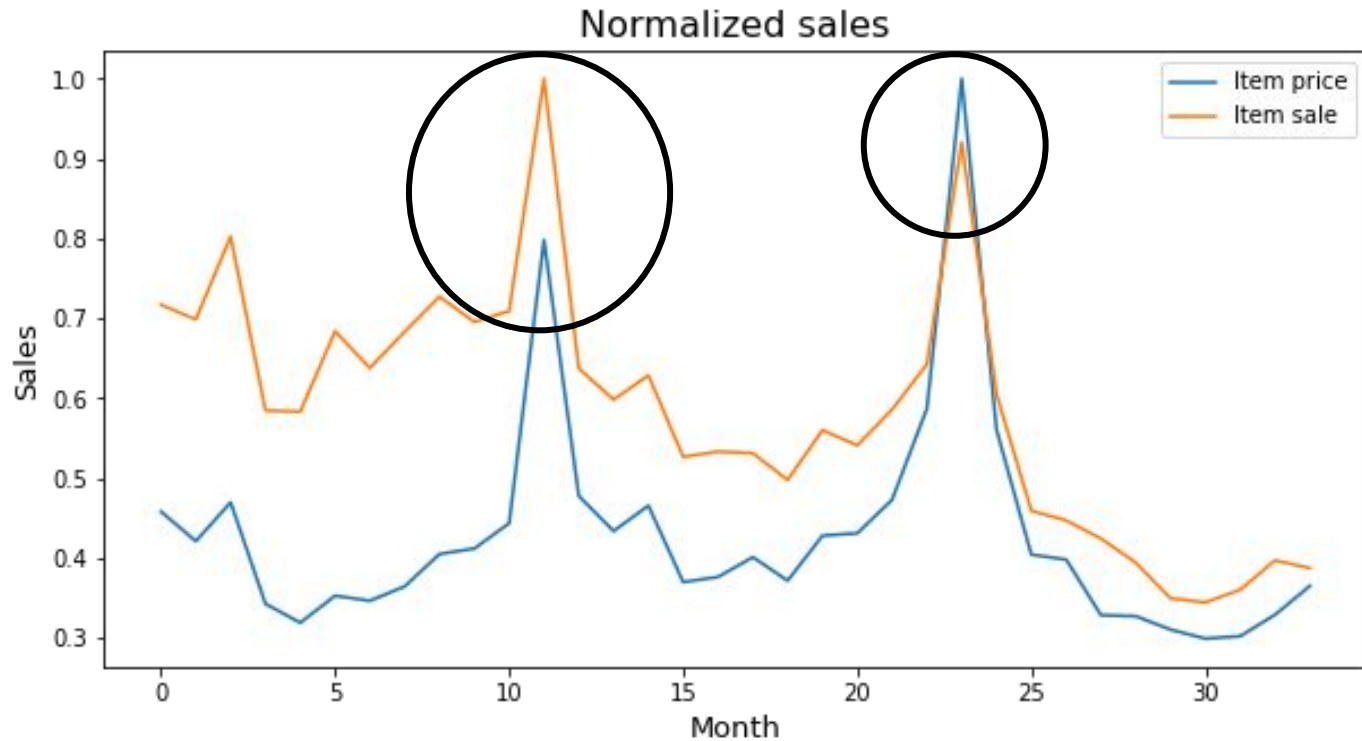
Month and sales impact



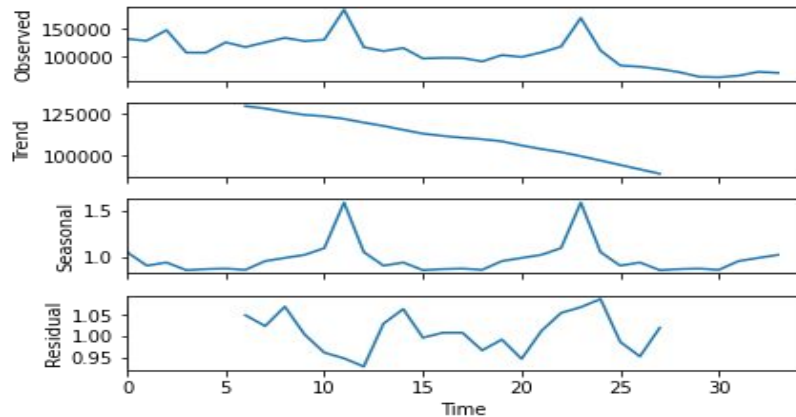
The plot of revenue in the total data



The plot of all items sold in the total data

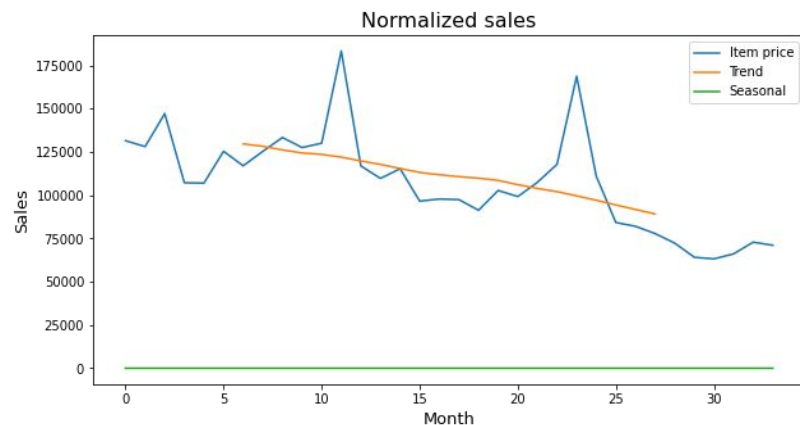


We can clearly see seasonality in the sales through the year, and correlation between the revenue and the item sales

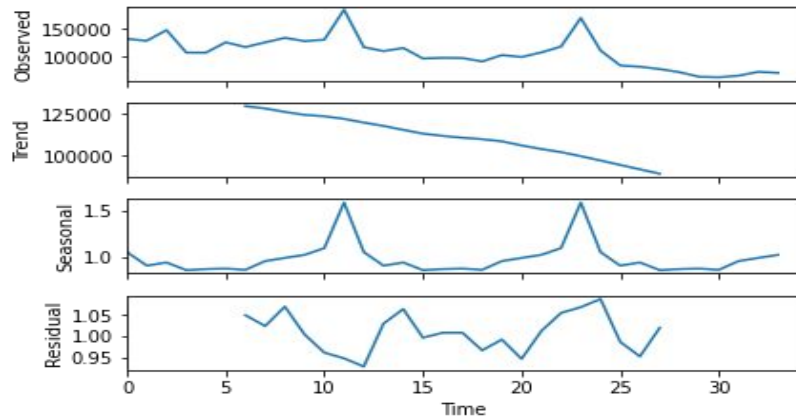


The seasonality is obvious, having peaks in certain months of the year. We will determine them.

The residual show us that there is a lot of variation in sales compared to the mean

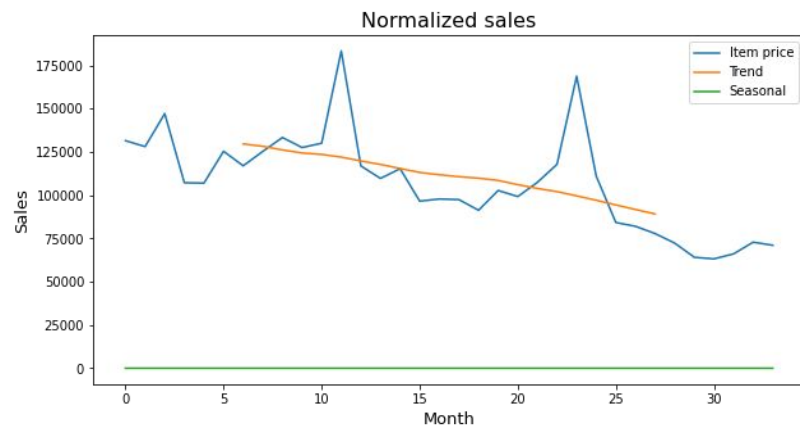


We can see that the trend goes down, so we can expect lower returns in the forecasting.



The seasonality is obvious, having peaks in certain months of the year. We will determine them.

The residual show us that there is a lot of variation in sales compared to the mean



We can see that the trend goes down, so we can expect lower returns in the forecasting.

- **Search for seasonality (estacionalidad)**
- **Which months are the most best for selling**
- **Which one the worst**
- **Which articles are the most sold per month**

**Day and the impact on
sales**

- **Same as month but for days**

**Does the GDP per
capita affects sales?**

- **Monthly impact and correlation for sales**

Forecasting sales

- **SVM algorithm**
 - **Features (input)**
 - **Month**
 - **Item**
 - **Store**
 - **Category**
 - **Forecasting (output)**
 - **Total monthly sales**

Conclusions