RETAIL: ASSESS SALES OUTLETS' PERFORMANCE

1/12/2016 – 2/28/2017

Conclusions and Recomendations

* In this analysis we studied how the outlets performed through the range of dates
* We found that a significant amount of data was missing in the customer_id column, at around 33%, which askewed all the data set toward Shop 0.
* Later we decided to removed all these rows as it seems that without them all the shops data was fairly uniform and meant that those missing rows were more like a fault in the data collection than an anomaly
* We found that a relative few customers have very big orders
* By looking for the best selling items we found that the top ones do not have a price per unit, this leads us to think again about the question of 'what was the best selling item' and instead ask 'what are the most profitable items' - this means that the most quantity sold doesn't mean they are the most valuable.
* And on that same line of thought we found that the shop with the most customers doesn't mean it is the shop with the best order's value
* Looking at overall shops performance we can say the value sold and customer count across the shops were quite uniform.
* And therefore I was begining to doubt the path of the idea on how to improve their performance
* Until we looked at the Loyalty membership which at first glance did not show any positive difference between those who were members and those who weren't.
* But by looking at the averages of value and orders quantity we found how effective the Loyalty program is in bringing value not to specifics shops but to the overall franchise
* Then we tested the Hypotheses about the proportions of the members of the loyalty program and value proportions of the members of the loyalty program and non-members
* The tests confirms that they were not equal and therefore we can say that the loyalty program does indeed affects the performance of all the shops(with a side note of the effect being positive)
* **Conclusions and Recommendation** :
    * We don't have enough data to advise on how to improve each shop per say as we don't know how and where they sale the products(geographically or online)
    * But by expanding the Loyalty program membership the revenue can be improved for all the shops (for all the franchise)
    * One recommendation would be to offer a Loyalty membership to those customers (not members) who tend to buy in bulk
    * Another one would be to lower the price of the membership in those shops with the lowest revenue
    * If we could have a detail customer profile we could target potentially better Loyalty members

# Overall look at the data set and date range

Events count by Date
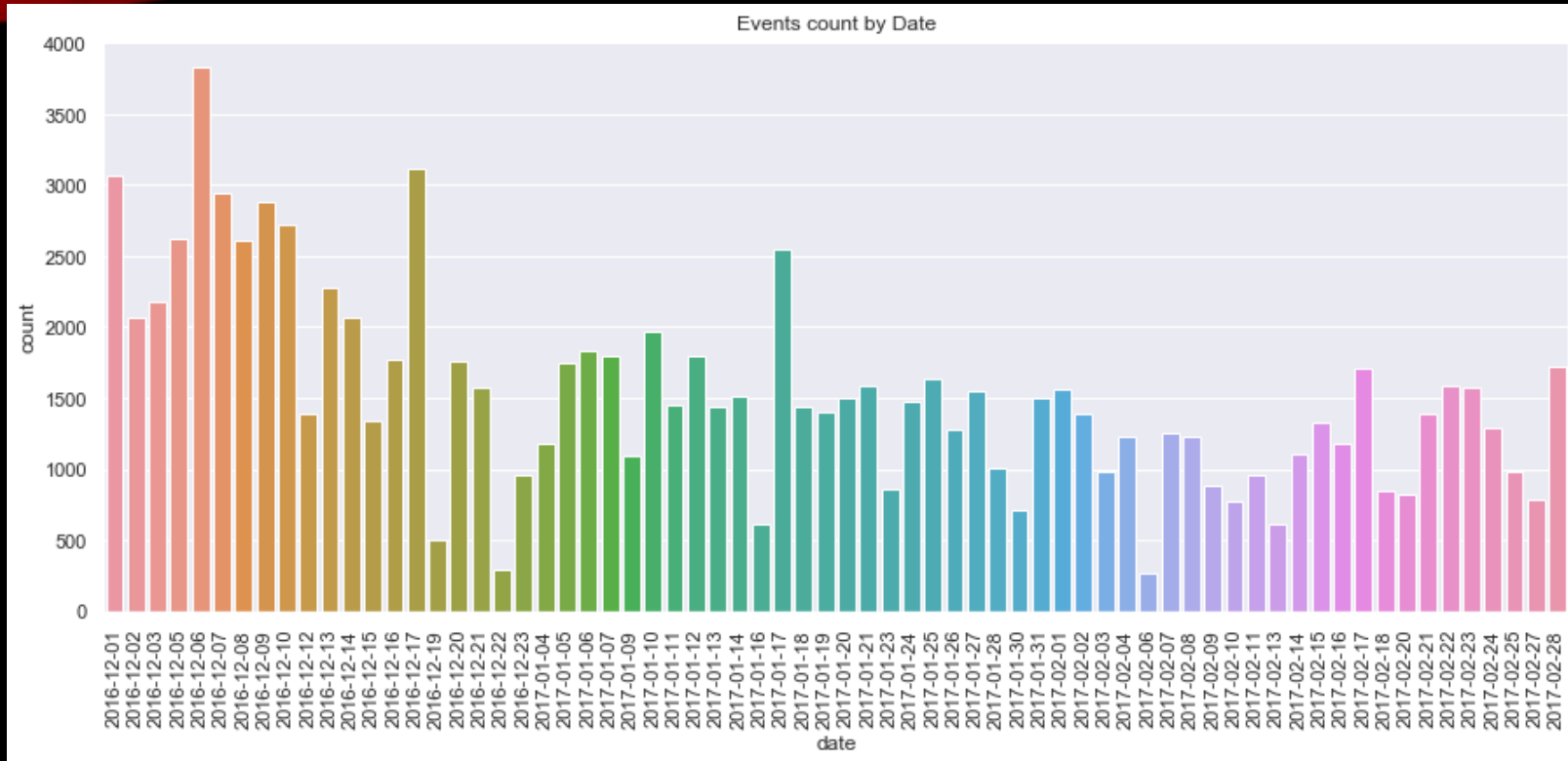
- Here we start with counting the purchases by date
- We try to identify some seasonality but we only see a significant peak at the beginning

We look at the purchases per customer

Number of purchases per customer

- We have a very skewed graph that tells us that a lot of clients order a few items and fewer clients order a lot of items

Value of orders

value_of_orders distributions
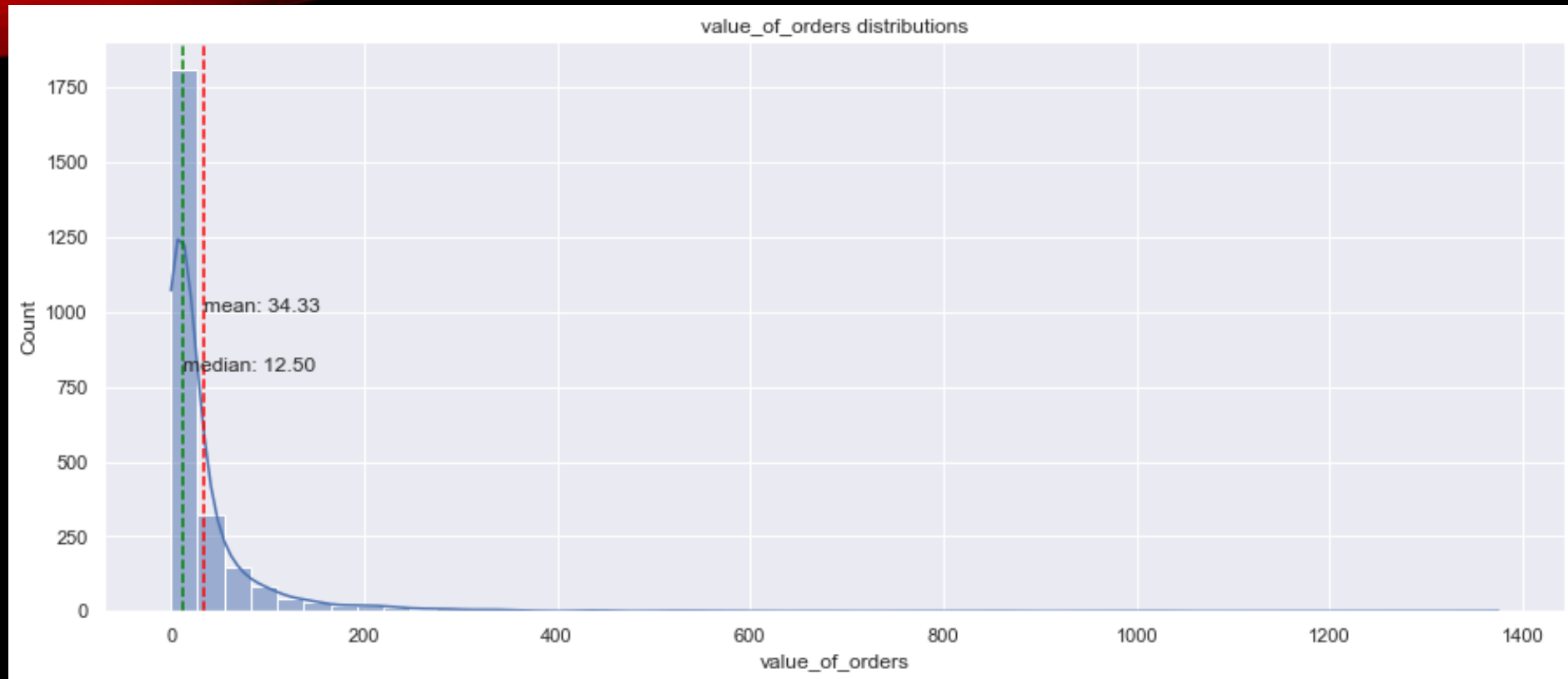
- Here we see that although most orders are at $34.33 there are very big orders
- This points out to bulk buyers

Items price distribution
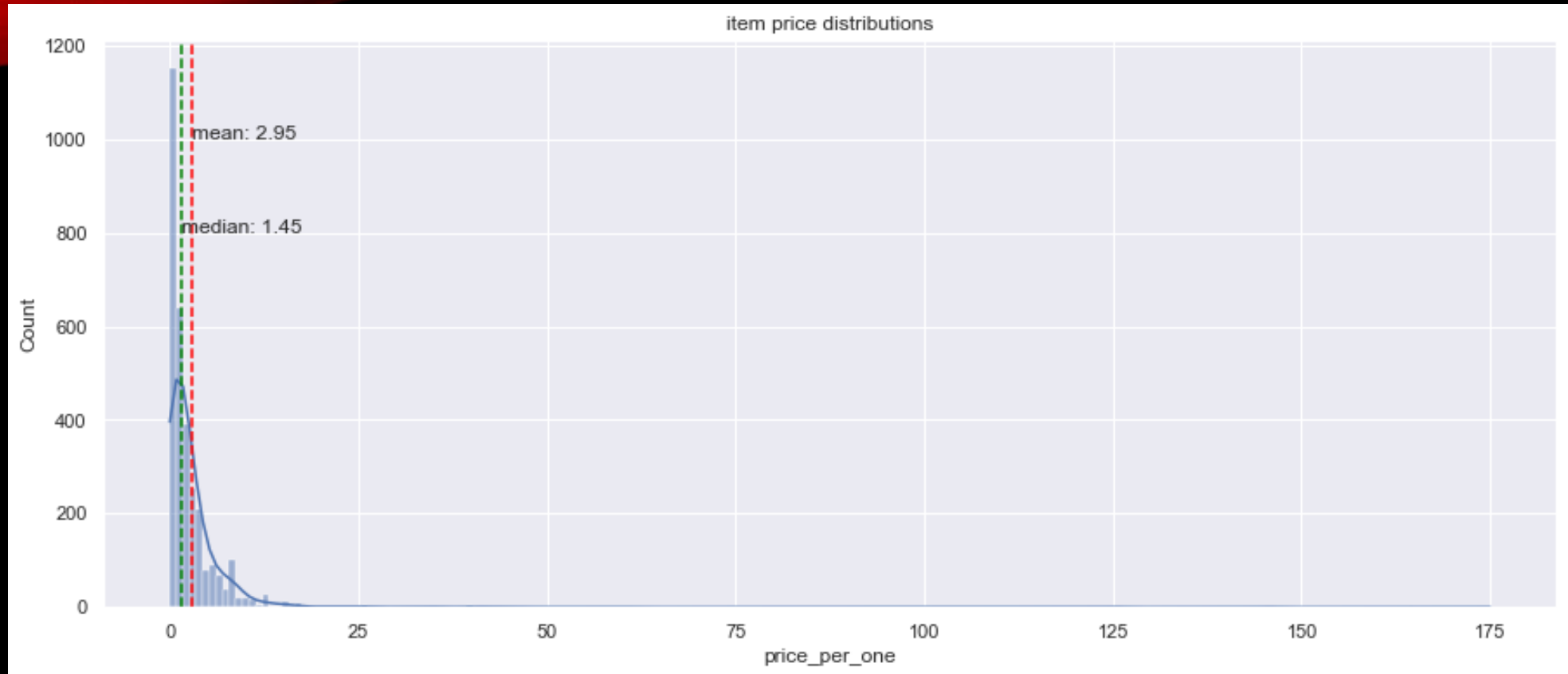
- Most item prices are very low and yet some are almost 100 times more expensive
- This graph makes the foundation to the claim that the most orders might not be the most valuable

Total order's value per Shop
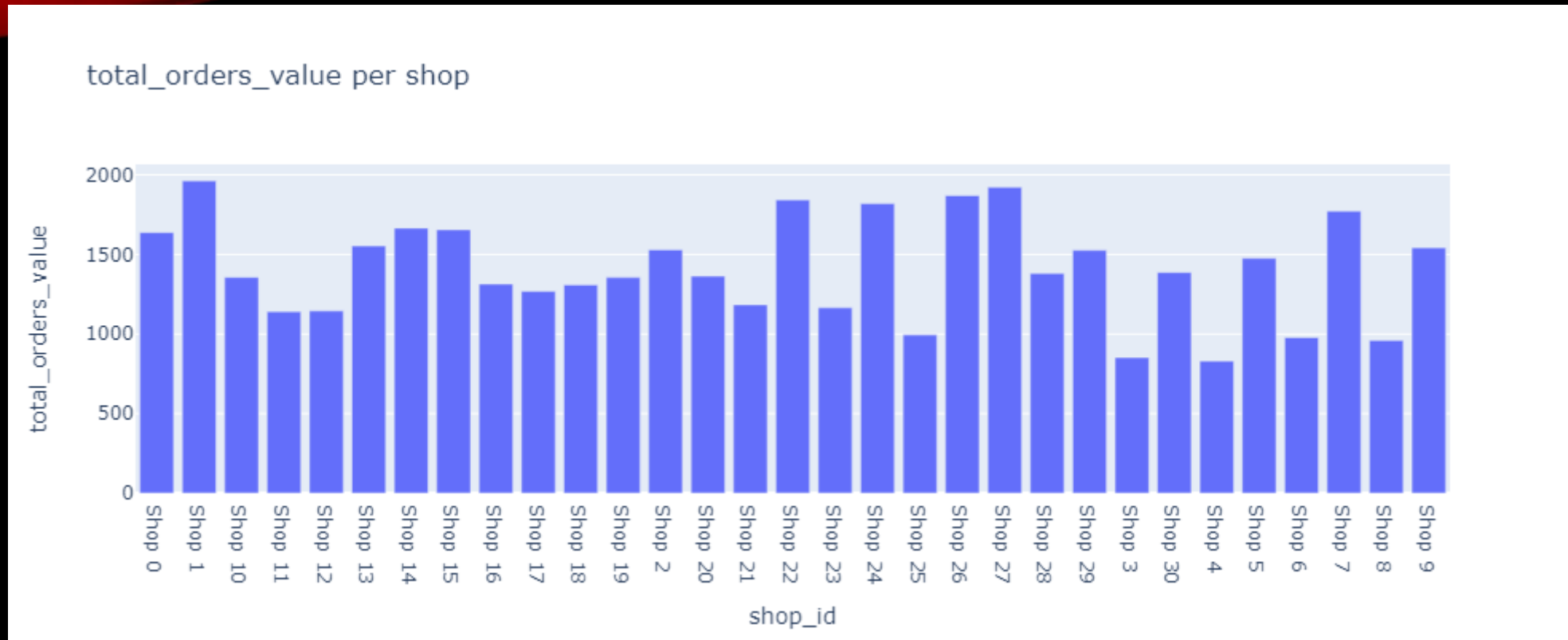
- Here we can see what value each shop brought in the time period we studied
- We can say that there are not any outstanding outliers and all shops performed fairly equal
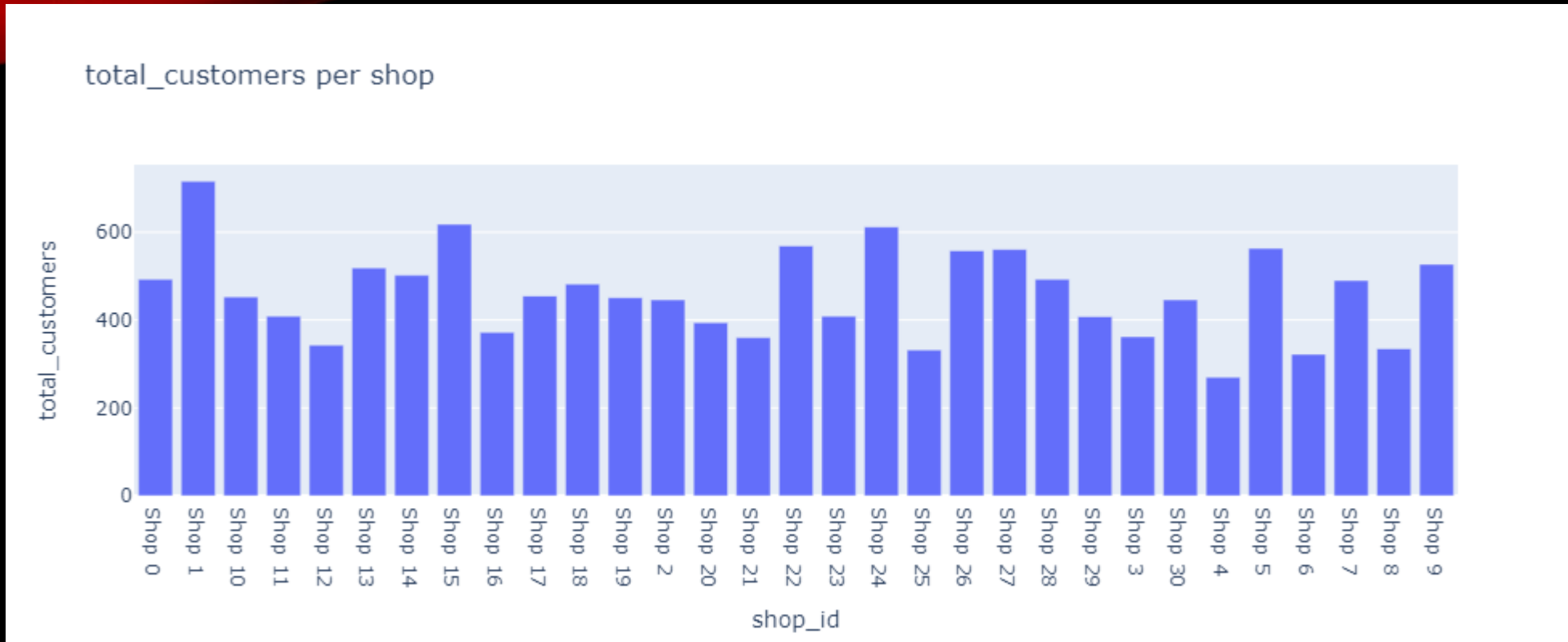
Total order's value per Shop

Total sales value per Shop distribution

mean: 1411.92

median: 1381.72
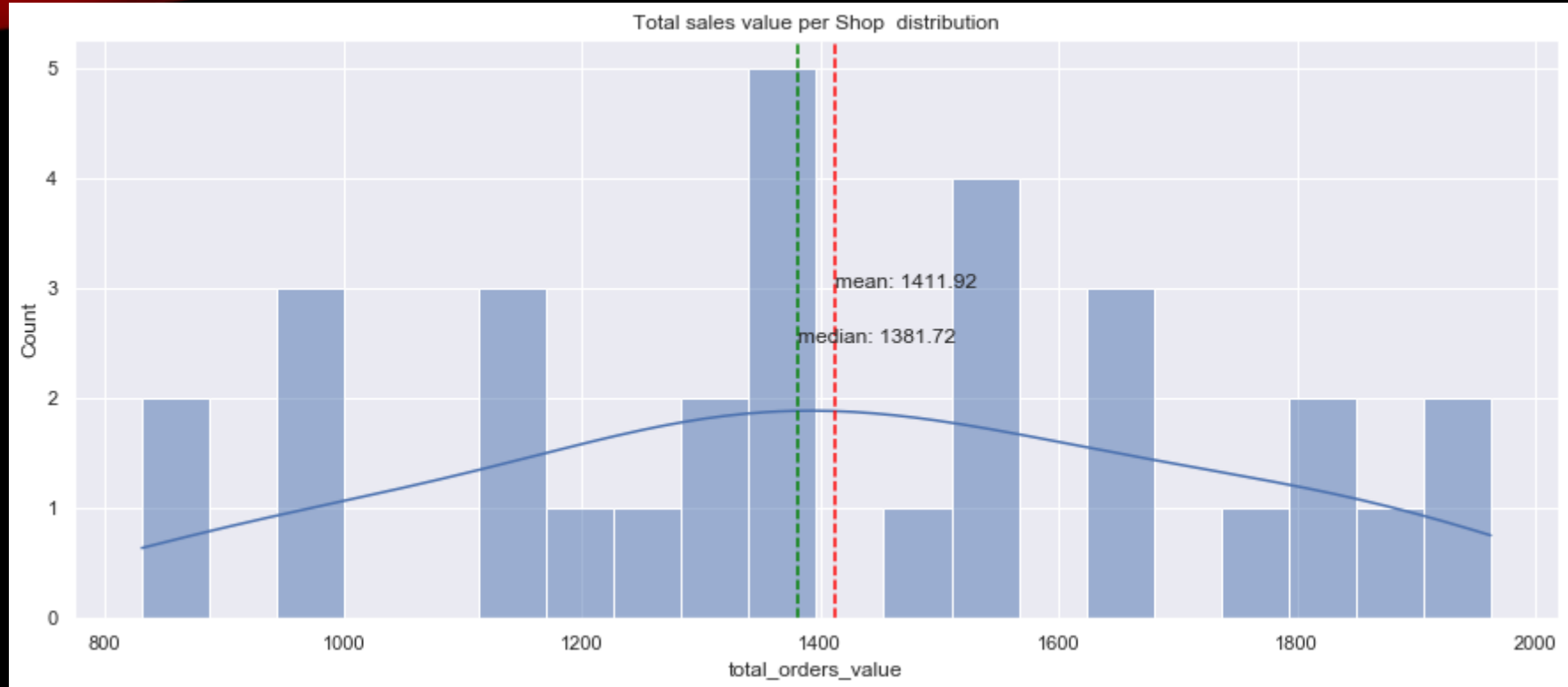
- To confirm the the claim the shops perform fairly equal we have an almost normal distribution of their order's value
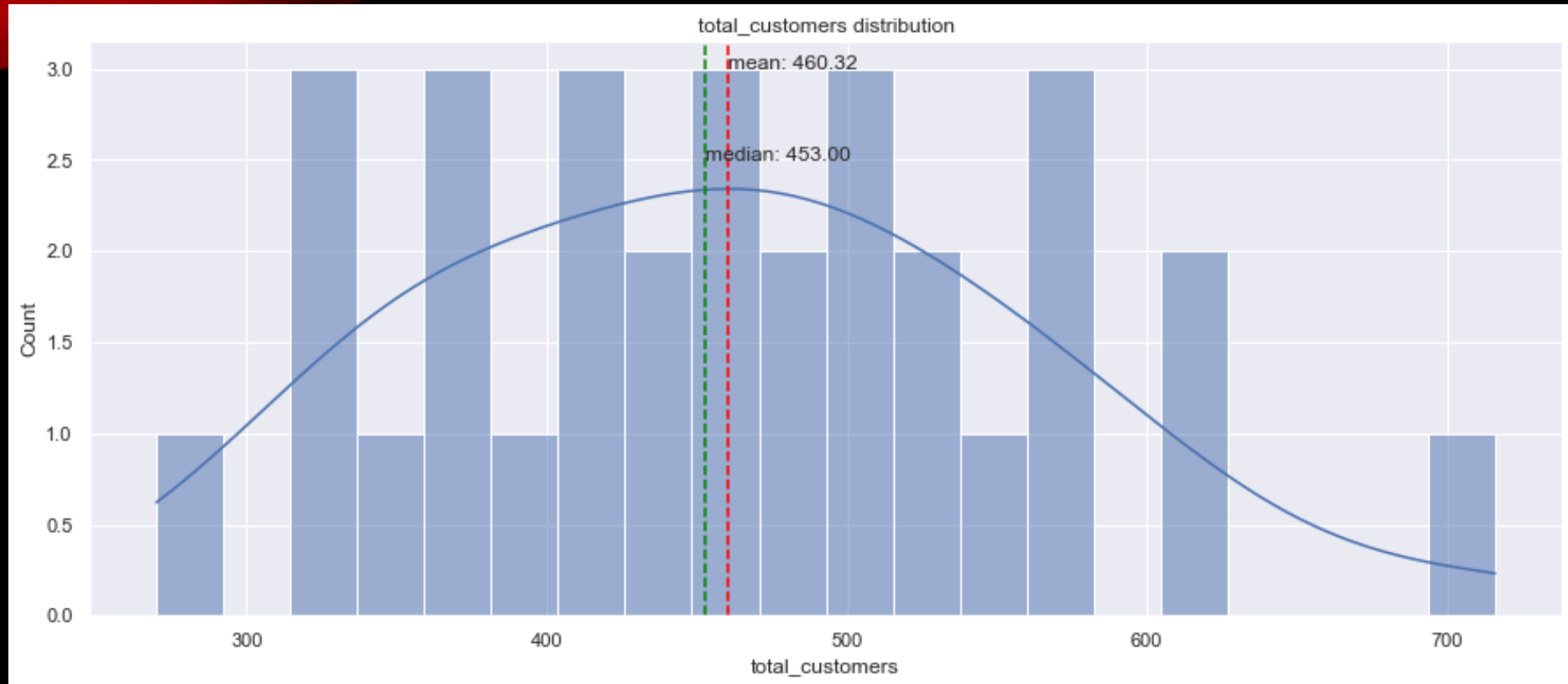
Total customers per Shop

- For total customers per shop we also identify a fairly equal distribution between shops

Total customers per Shop

- In similar manner as the value per shop we have a almos normal distribution
- Here I though I was going to identify a set of shops that were not performing as the others

Loyalty Members vs Non Members

Who buys more by quantity

loyalty_program	total_value_purchased	total_purchases_count
0	26670.10	8245
1	17099.37	6025

- By looking at all the set filtered by loyalty membership we discover that Members do bring relative more value
- This could mean that if the membership pool is expanded the franchise as a whole could bring more revenue

Who buys more by MEAN

loyalty_program	MEAN_value_purchased	total_purchases_count
0	3.23470	8245
1	2.83807	6025

Who buys more by MEDIAN

loyalty_program	MEDIAN_value_purchased	total_purchases_count
0	1.65	8245
1	1.65	6025

Loyalty Members vs Non Members

Top 5 shops purchase values avg by membership and non membership

	members_top_5_avg	non_members_top_5_avg
0	145.0	145.000
1	125.0	145.000
2	125.0	125.000
3	125.0	76.750
4	125.0	65.975

Top 5 shops purchase count avg by membership and non membership

	members_top5_quantity	non_members_top5_quantity
0	131	98
1	104	82
2	101	81
3	80	73
4	68	72

```
1  value_comparison_df.mean()
members_top_5_avg        129.000
non_members_top_5_avg    111.545
```

```
1  quantity_comparison_df.mean()
members_top5_quantity        96.8
non_members_top5_quantity    81.2
```

- To make sure my hunch about the loyalty membership I took a snap at the top 5 avgs for purchase value and the top 5 quantity count
- The tables at least do not contradict this vector of analysis
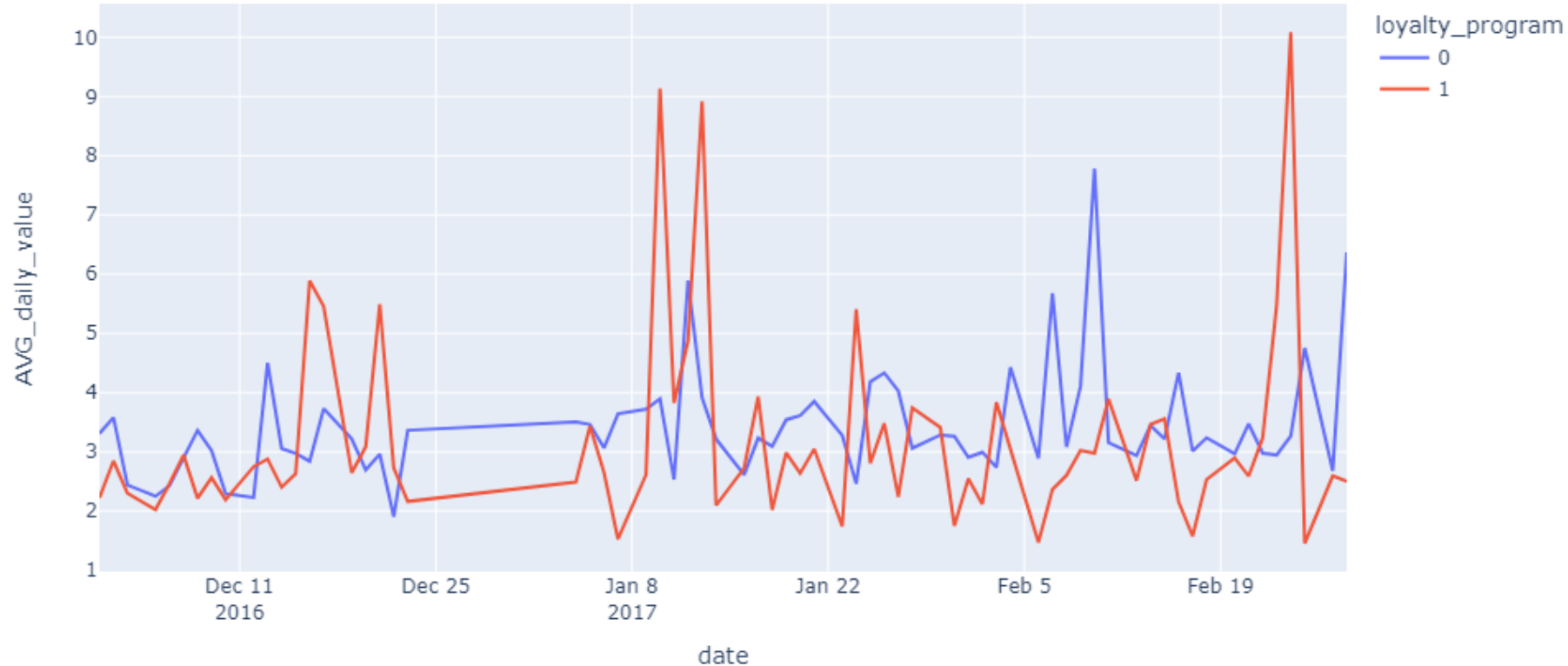
Loyalty Members vs Non Members

- And here we have the summary for all the set by loyalty membership
- If we only look at purchases per day our intuition points ot the non members bringing it the most revenue

Loyalty Members vs Non Members

AVG daily_value per day by loyalty program membership

- But If we only look at AVG purchases per day we see how Loyalty members tend to spend much more than non members from time to time and therefore the conclusion that the Loyalty program is the key to improve overall performance