

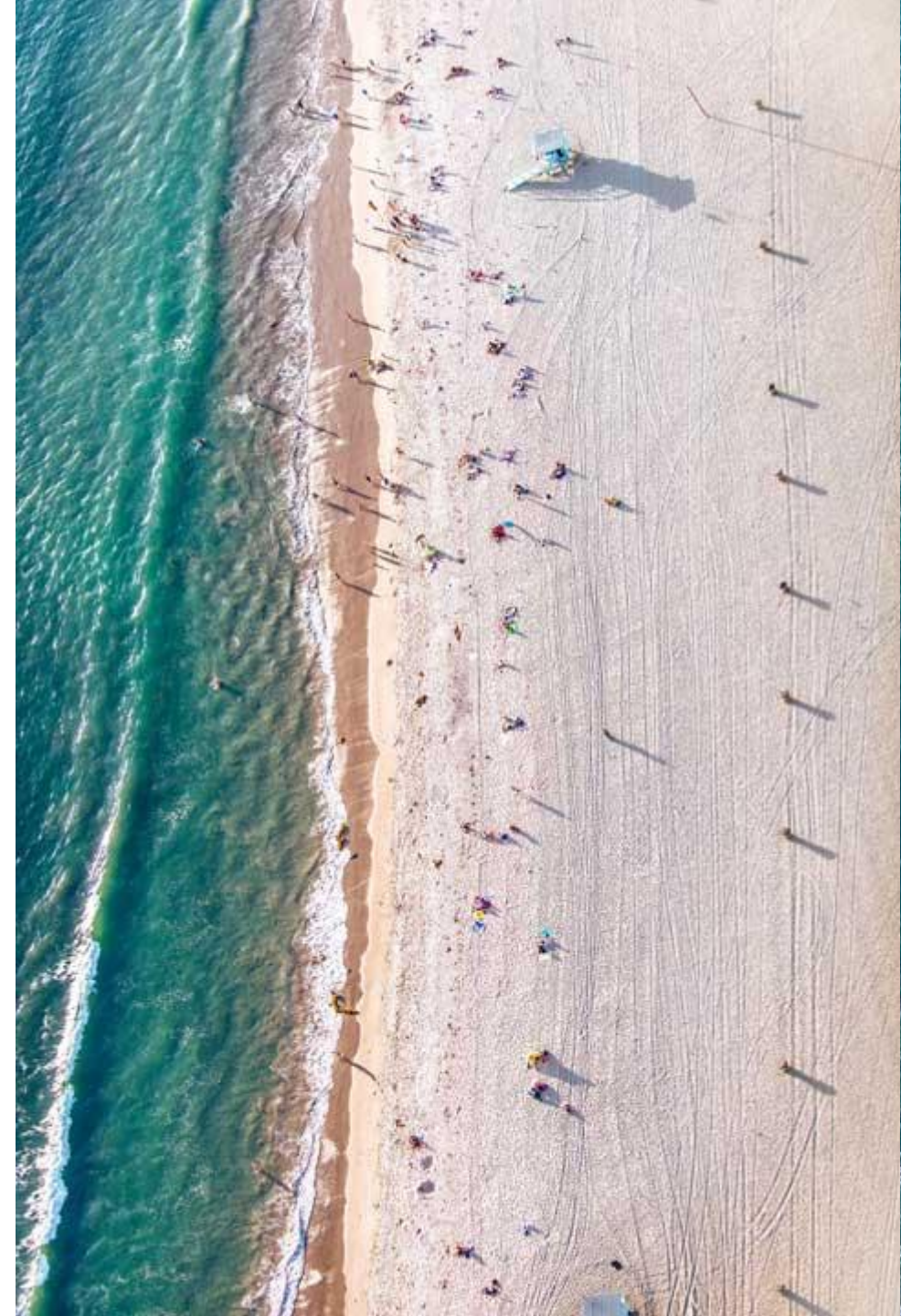
June 2020

Category review: Chips

Retail Analytics



Classification: Confidential



Our 17 year history assures best practice in privacy, security and the ethical use of data

We all have a responsibility to use data for good

Privacy

- We have built our business based on privacy by design principles for the past 17 years
- Quantum has strict protocols around the receipt and storage of personal information
- All information is de-identified using an irreversible tokenisation process with no ability to re-identify individuals.

Security

- We are ISO27001 certified - internationally recognised for our ability to uphold best practice standards across information security
- We use 'bank grade' security to store and process our data
- Comply with 200+ security requirements from NAB, Woolworths and other data partners
- All partner data is held in separate restricted environments
- All access to partner data is limited to essential staff only
- Security environment and processes regularly audited by our data partners.

Ethical use of data

Applies to all facets of our work, from the initiatives we take on, the information we use and how our solutions impact individuals, organisations and society.

Quantum believes in using data for progress, with great care and responsibility. As such please respect the confidence nature of this document.

Executive summary

01

Task 1

Data Preparation and Customer Analysis

1. Creating and interpreting high level summaries of the data.
2. Finding outliers and removing them.
3. Checking data formats and correcting it.

02

Task 2

Experimentation and Uplift Testing

1. Total sales revenue
2. Total number of customers
3. Average number of transactions per customer

01

Data Preparation and Customer Analysis

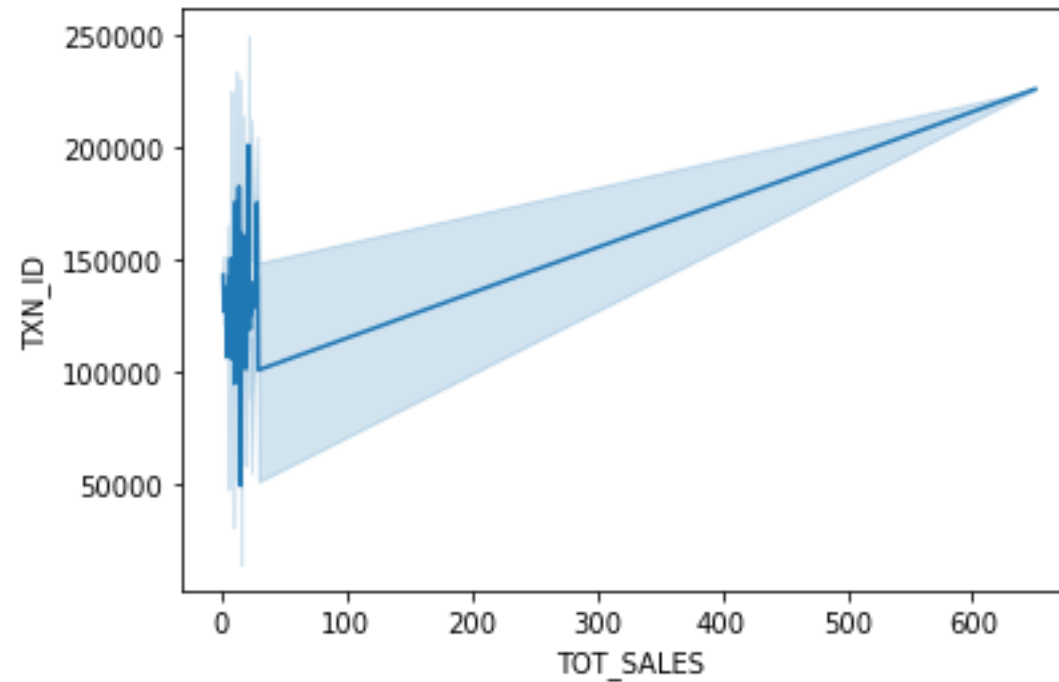
Here we first Analysis the data set Using Python

We check for any null or missing value in dataset.

Also saw the summary of data provided using Describe method in python

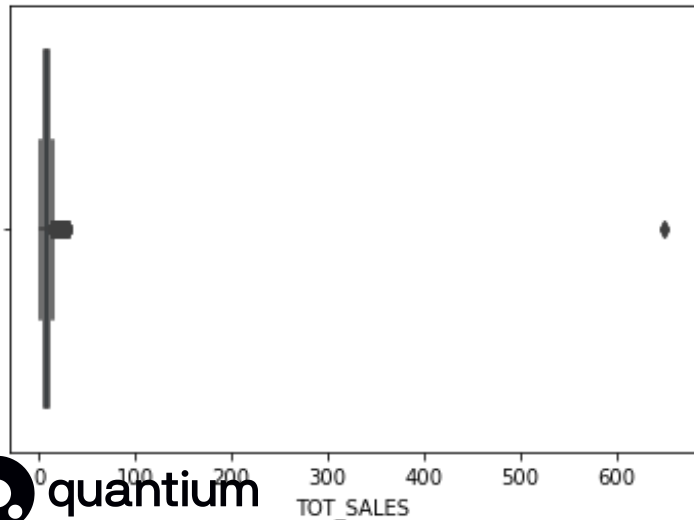
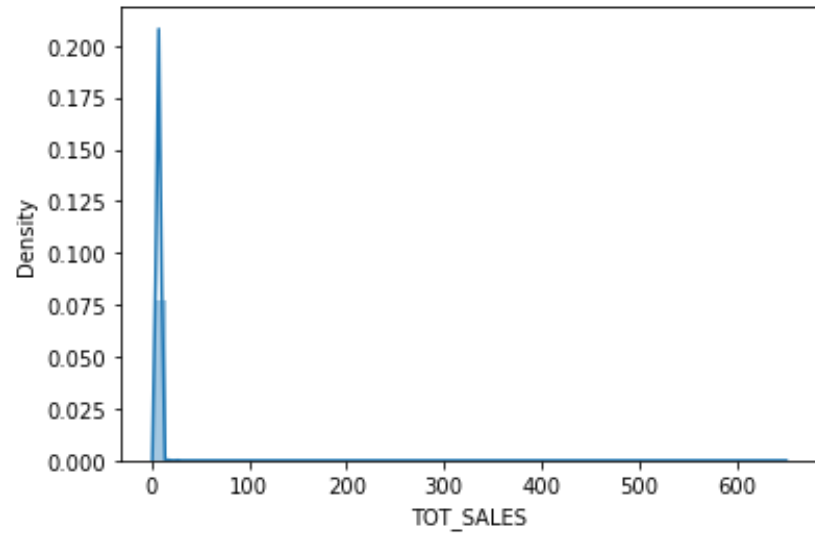
	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR
count	264836.000000	264836.00000	2.648360e+05	2.648360e+05	264836.000000
mean	43464.036260	135.08011	1.355495e+05	1.351583e+05	56.583157
std	105.389282	76.78418	8.057998e+04	7.813303e+04	32.826638
min	43282.000000	1.00000	1.000000e+03	1.000000e+00	1.000000
25%	43373.000000	70.00000	7.002100e+04	6.760150e+04	28.000000
50%	43464.000000	130.00000	1.303575e+05	1.351375e+05	56.000000
75%	43555.000000	203.00000	2.030942e+05	2.027012e+05	85.000000
max	43646.000000	272.00000	2.373711e+06	2.415841e+06	114.000000

We saw the line graph between two main Index
Total Sales and Taxation Id of Customer.

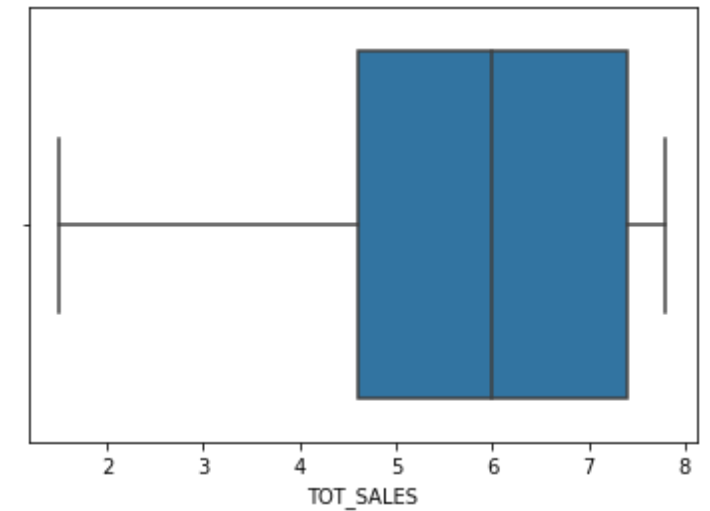
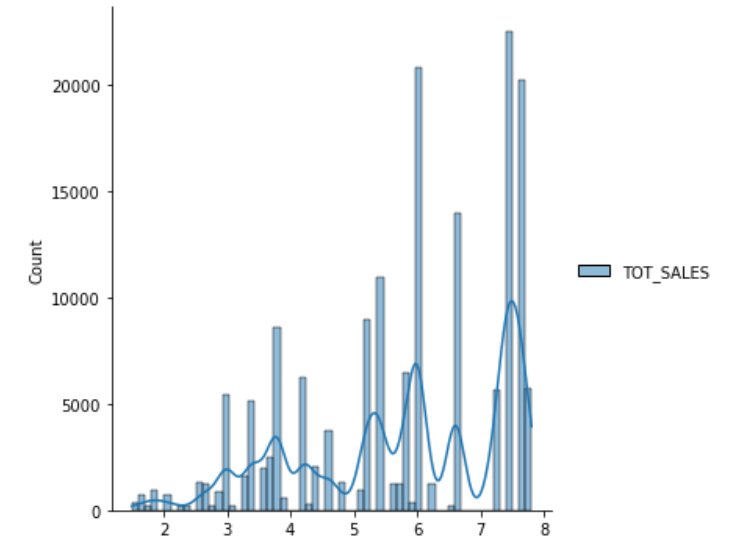


This Boxplot of dataset Vs Distplot of dataset

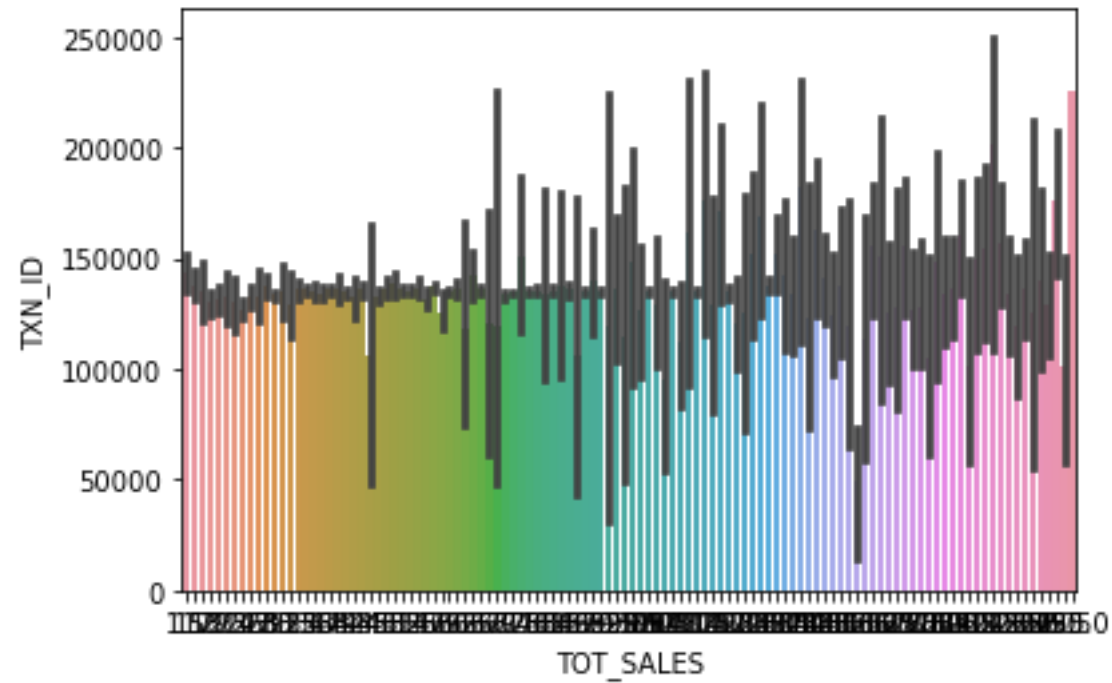
With Outliers



After Removing Outliers



Bar plot for the dataset provided



02

Experimentation and Uplift Testing

In this we calculated Total Sales Revenue

And Total Customer Id which is TXN_ID in this data set.

Total Sales Revenue

```
total_sales = sum(dataset['TOT_SALES'])  
print(total_sales)
```

```
[5]  
... 1933114.9999996515
```

There is not any customer data available for this project. But we can get customer by TXN_ID because it is unique for each individual customer.

```
dataset.describe()
```

	LYLTY_CARD_NBR	STORE_NBR	TXN_ID	PROD_NBR	PROD_QTY	TOT_SALES	PACK_SIZE
count	2.648340e+05	264834.000000	2.648340e+05	264834.000000	264834.000000	264834.000000	264834.000000
mean	1.355488e+05	135.079423	1.351576e+05	56.583554	1.905813	7.299346	182.425512
std	8.057990e+04	76.784063	7.813292e+04	32.826444	0.343436	2.527241	64.325148
min	1.000000e+03	1.000000	1.000000e+00	1.000000	1.000000	1.500000	70.000000
25%	7.002100e+04	70.000000	6.760050e+04	28.000000	2.000000	5.400000	150.000000
50%	1.303570e+05	130.000000	1.351365e+05	56.000000	2.000000	7.400000	170.000000
75%	2.030940e+05	203.000000	2.026998e+05	85.000000	2.000000	9.200000	175.000000
max	2.373711e+06	272.000000	2.415841e+06	114.000000	5.000000	29.500000	380.000000

At Last we calculated total Number of Transaction Per Customer.

```
total_customer = max(dataset['TXN_ID'])  
print(total_customer)
```

2415841

TOTAL NO. OF TRANSACTIONS PER CUSTOMER

```
dataset.shape
```

(264834, 12)

```
avg_trans = (total_customer)/264834  
print(avg_trans)
```

0.9122095350294902

Thank You

Disclaimer: This document comprises, and is the subject of intellectual property (including copyright) and confidentiality rights of one or multiple owners, including The Quantum Group Pty Limited and its affiliates (**Quantium**) and where applicable, its third-party data owners (**Data Providers**), together (**IP Owners**). The information contained in this document may have been prepared using raw data owned by the Data Providers. The Data Providers have not been involved in the analysis of the raw data, the preparation of, or the information contained in the document. The IP Owners do not make any representation (express or implied), nor give any guarantee or warranty in relation to the accuracy, completeness or appropriateness of the raw data, nor the analysis contained in this document. None of the IP Owners will have any liability for any use or disclosure by the recipient of any information contained in, or derived from this document. To the maximum extent permitted by law, the IP Owners expressly disclaim, take no responsibility for and have no liability for the preparation, contents, accuracy or completeness of this document, nor the analysis on which it is based. This document is provided in confidence, may only be used for the purpose provided, and may not be copied, reproduced, distributed, disclosed or made available to a third party in any way except strictly in accordance with the applicable written terms and conditions between you and Quantum, or otherwise with Quantum's prior written permission