

Captain Book

A SALES ANALYSIS

Table of Content

- Project Purpose
- Presentation of Data
- Monthly Sales
- **Meet your demographic** (Univariate)
 - Univariate Analysis
 - Concentration Analysis
- **Customer Spending Habits** (Bivariate)
- **Product Category Analysis** (Bivariate)
- **Misc Correlation Analysis** (Bivariate)
- Results

Purpose of Report

- **Analyse the sales of the company**
 - Target new areas of growth
- **Discover and Clean Data**
- **Univariate Analysis**
 - Central Tendency and Dispersion
 - Concentration Analysis
 - Graphic Representation of data
- **Bivariate Analysis (Correlation)**
 - Gender and product Category
 - Age vs (Spending, Purchase Frequency, Basket Size, Category)

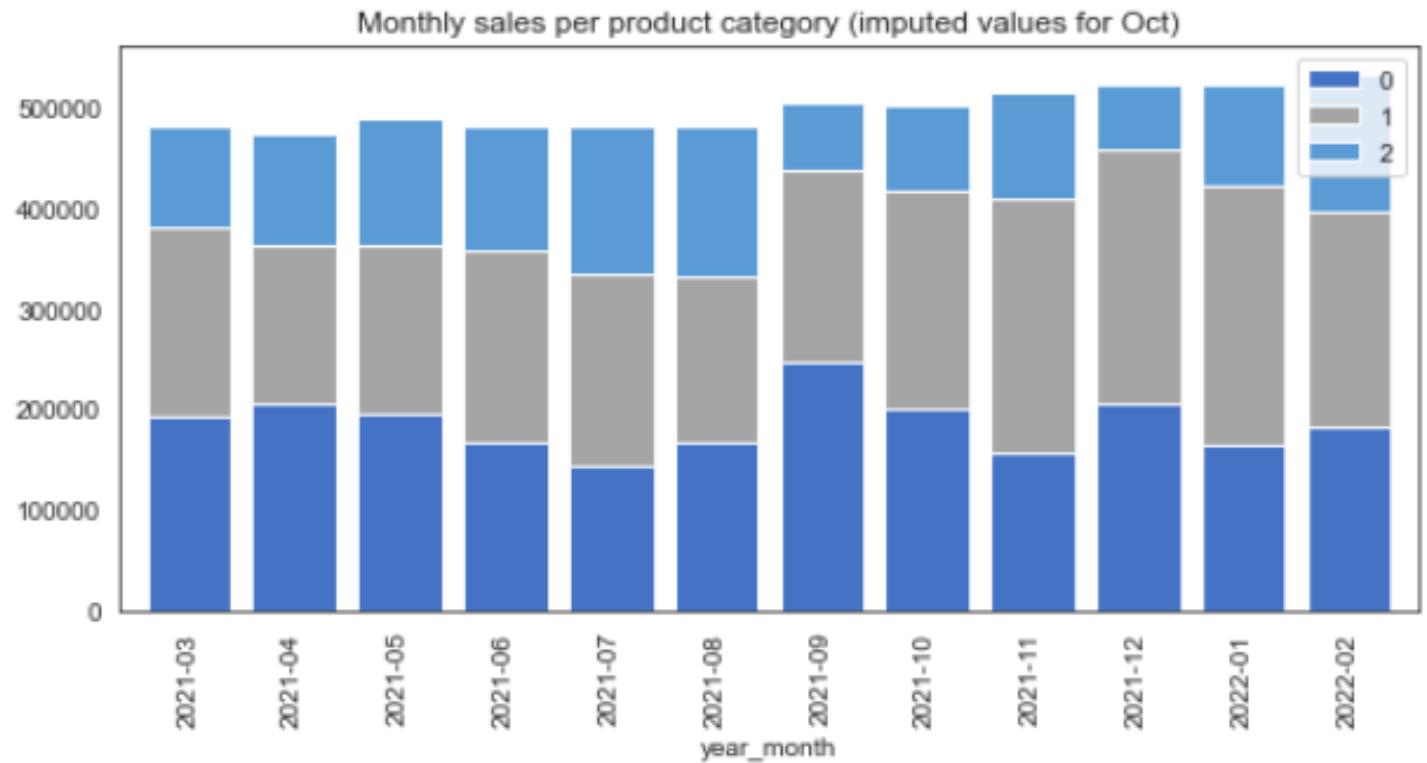
Data discovery and cleaning

- **We used three tables of internal data:** transactions, customers and products
- **Data is statistically significant** (337k transactions)
 - Ranged from 2021-03-01 to 2022-02-28
 - Covers 3,287 product in 3 product categories
 - Spread over 8623 customers
- **Data Cleaning**
 - 200 Test Orders in the transaction table (**removed**)
 - Missing transactions found (Oct 2-27 2021) for product category 1 (**imputation**)
 - Outliers, 23k transactions from 4 customers (**removed**)
 - 103 transactions found without corresponding product (**removed**)

Monthly Sales

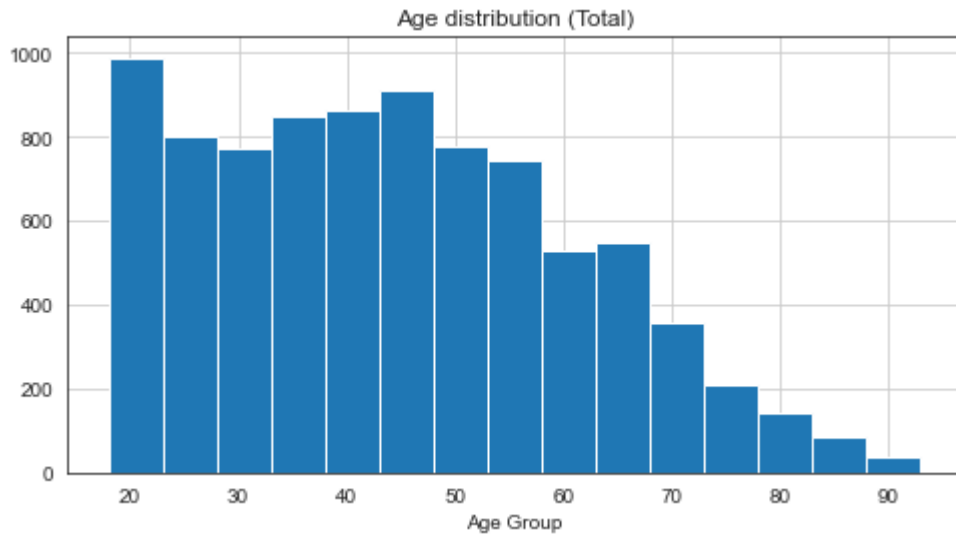
From Mar 2021 to Feb 2022,
by category

- Monthly Sales: 498k/mo USD
- Top product: Cat 1
202k/mo USD
- Growth Rate: 0.98%/mo

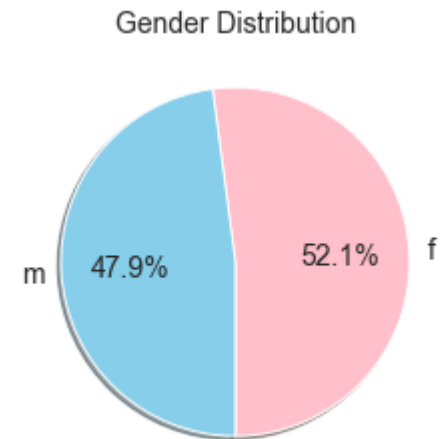


Meet your demographic

We are interested in behaviour predicated on customer age and gender



Average Age: 43.7 (mode 18)
Standard Dev: 16.9
Skew: 0.36 (positively skewed)
Kurtosis: -0.63 (flatter than a normal distrib)



Sex	Count	Ratio
Female	4,479	52.1%
Male	4,121	47.9%

Results

We have a wide spread of ages (kurtosis) with a slight tendency of younger customers (pos skew)

We see an even gender distribution with a small percentage point Towards having more females

Concentration Analysis

Are sales distributed evenly per customer?



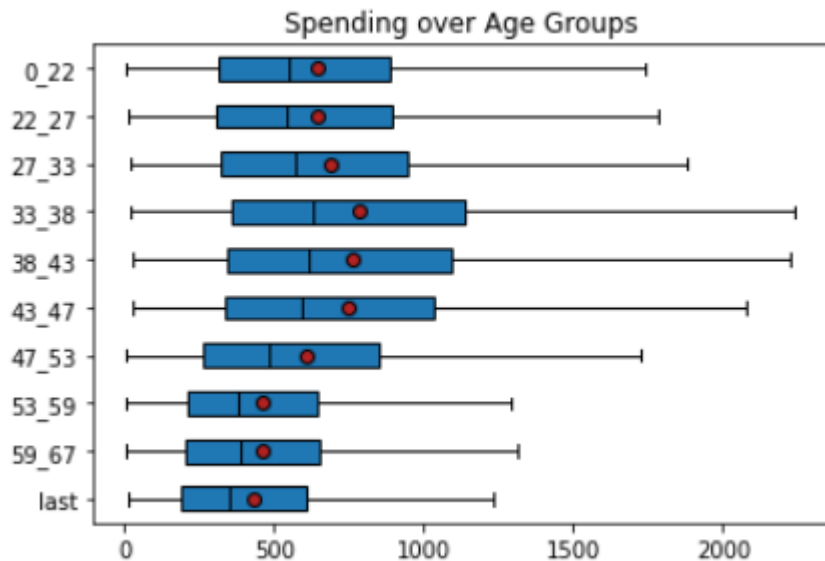
Lorenz Curve

Gini Index: 0.395

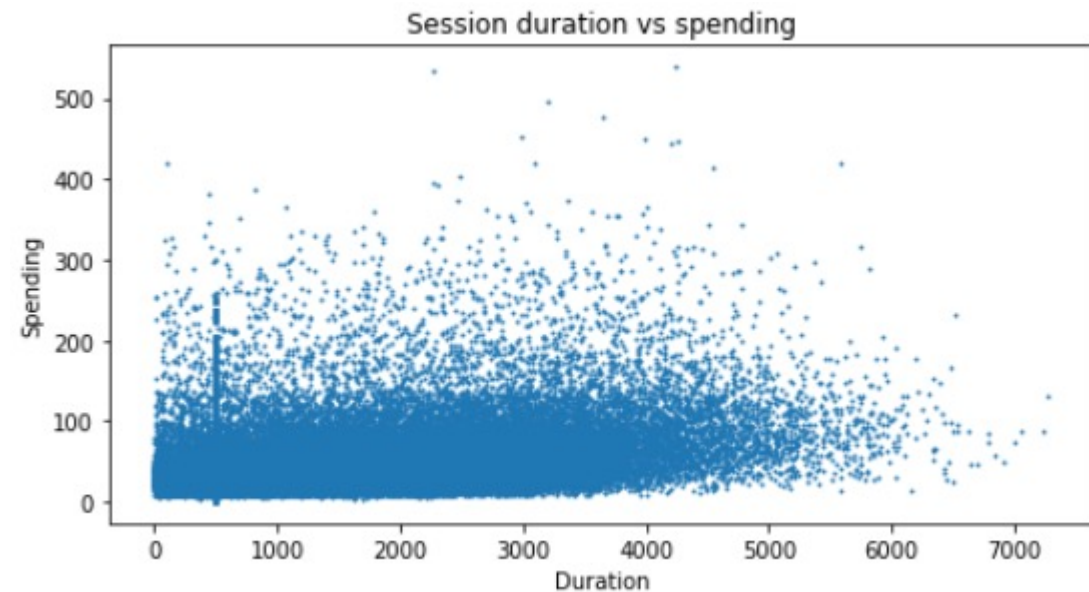
- Our Lorenz curve show that 20% of the customers are responsible for almost half (44%) of the sales
- Correcting our data for outliers reduces our Gini Index from 0.440 to 0.395

Customer Spending Habits

Can we correlate age and/or gender with spending?



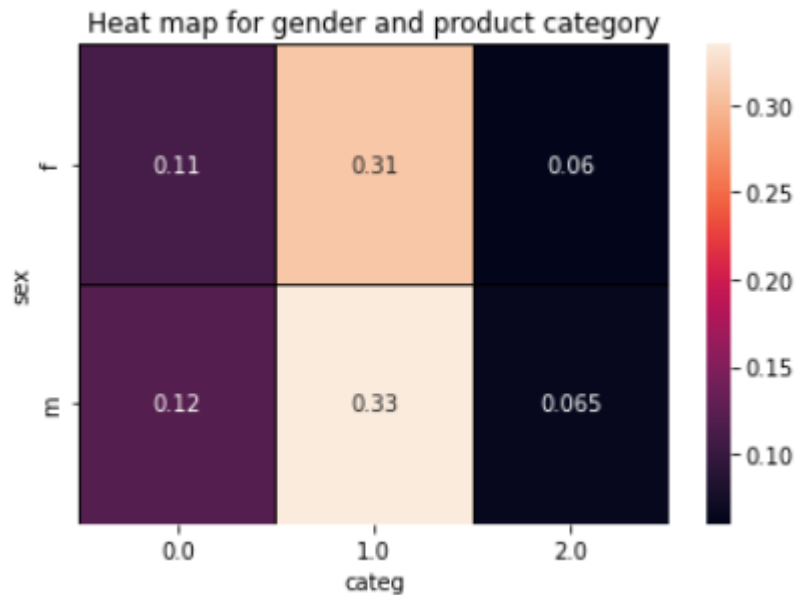
- Correlating Age Group (qualitative) and Spending (quantitative)
- **Eta-squared** correlation: 0.07 (low)
- Biggest spenders 33-38 and 38-43 year olds



- Correlating Spending (quantitative) and Session Duration (quantitative)
- **Pearson Coefficient:** 0.45 (moderate)
- We've found a correlation between session duration and spending using linear regression

Product Category Analysis

Part 1: Which products should we market to whom? Product **Category** and **Gender**



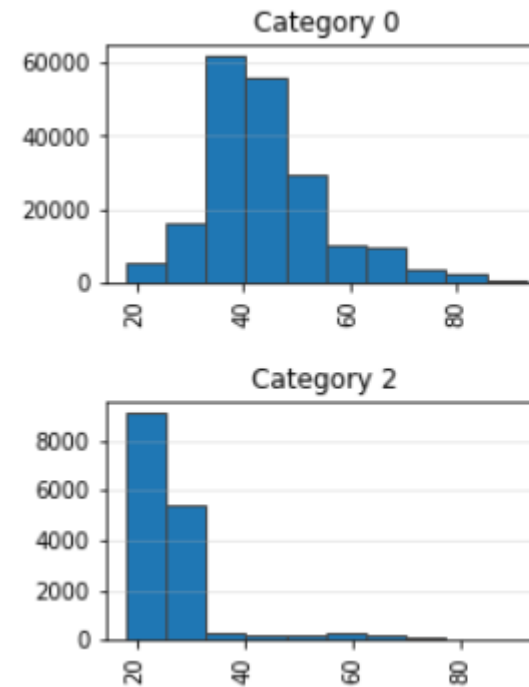
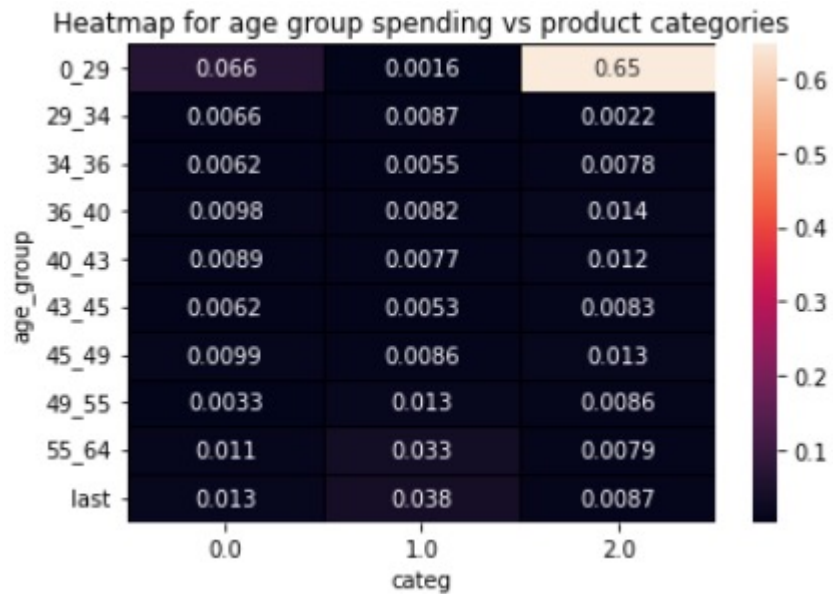
- Correlating gender (qualitative) and product category (qualitative)
- We use the Chi-square method to plot a heatmap
 - Chi-value: **10.2**

Results

- the highest values are around product category 0.3 and that this goes for both genders. This indicates that product category 1 is most responsible for non-independency across both genders
- The xi-value is 10.2
- Excluding extreme values

Product Category Analysis

Part 2: Which products should we market to whom? Product **Category** and **Age**



Plotting spending over age groups for each product category highlights what our heat map indicated more clearly.

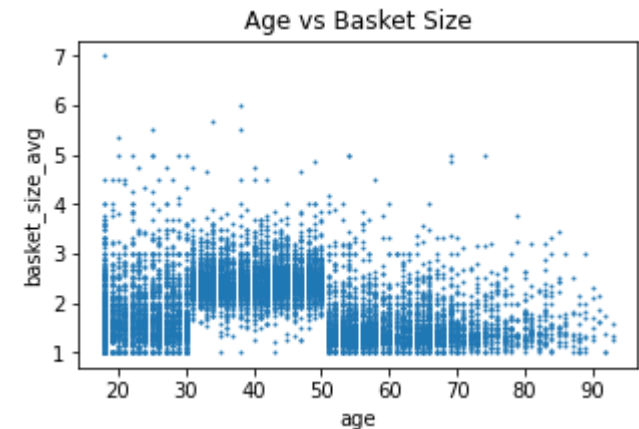
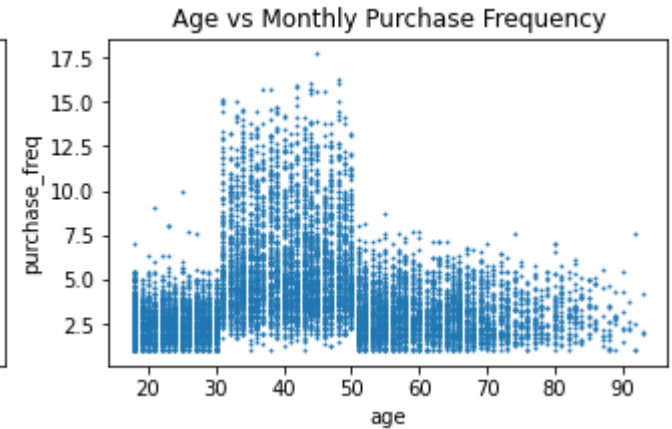
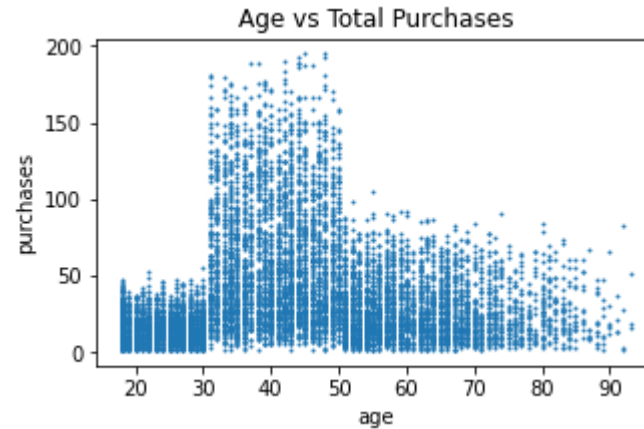
Product Category 2 almost exclusively caters for the younger clients

- We correlate Age group (qualitative) with product category (qualitative)
- Using the chi-square method again, we find that product Category 2 deviates from the expected quite a lot.

Misc Correlations

Part 3: How does age correlate with purchase data?

- We also tried to correlate age (quantitative) to the following
 - Total amount of purchases (quantitative)
 - R-Square method: 0.03 Weak Correlation
 - Monthly Purchase Frequency (quantitative)
 - R-Square method: -0.03 Weak Correlation
 - Basket Size (quantitative)
 - R-Square method: -0.23
- We do not find good age-related linear correlations here
 - Using scatter-plot and r-square is not working well
- We do find higher spending habits among 31-50 years old
 - Similar to slide-8 “Customer Spending Habits”



Conclusions

- Our average customer is on average 44 years old
 - Slightly leaning towards the younger side of the age distribution
 - 2% more women than men buying our products
- There is no gender bias for any product
- 20% of our customers are responsible for 44% of the purchases
 - Number goes up to 48% if we include outliers
- Category 2 products are almost exclusively bought by people < 30 years old
- Age group 31-50 is spending more money, making purchases more often and buy more products