FAIR FOOD DESCRIPTIVE ANALYSIS

Statistical Thinking and analysis

Lim Boon Yan Edward

Edward.lim.2025@mitb.smu.edu.sg

PROBLEM STATEMENT

FairFood aims to analyze the performance of its product categories up to 30 June 2024 and explore new opportunities to further grow its top-performing category.

Analytical Question 1 (Descriptive):
Which product category contributes the most to overall sales, and within that category, which demographic group represents the largest share of purchases?

- **Aim:** To identify top volume driver category and the demographics involve.
- **Purpose:** To determine which product category drives the highest sales volume and the demographic segments contributing to it.
- **Benefit:** Provides guidance to FairFood in optimizing resource allocation to enhance customer targeting.

Analytical Question 2 (Inferential):
Is there a significant difference in purchasing behavior among the marital status groups?

- **Aim:** To understand whether marital status affects wine purchases.
- **Purpose:** To assess, through inferential testing, whether differences in wine purchase patterns across marital groups are supported by statistical evidence rather than random variation.
- **Benefit:** The findings will help FairFood make more informed decisions regarding marketing investments and customer segmentation strategies, ensuring resources are directed toward groups with meaningful purchasing differences.

DATA PREPARATION

In preparation for analysis the following cleaning have been done are as illustrated in the below table:

Change of Data Type, Creation of new columns, Dropping of columns, Dropping of Rows

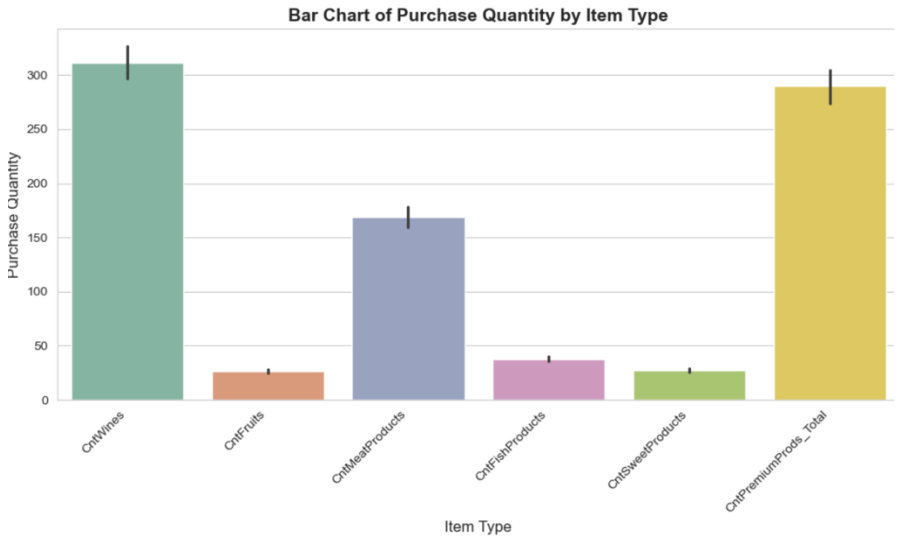
Change of Data Type	
Data Column	**Casting data type**
EnrolmentDate	Date Time
Education	Category
Marital_status	Category

Columns Created	
Column Name	**Formula**
Total_kids	kidhome + teenhome
purchase_freq	NumWebPurchases + NumAppPurchases + NumStorePurchases
CntPremiumProds_Total	CntPremiumProds + CntPremiumProds.1
Total_product_purchases	CntWines +CntFruits + CntMeatProducts + CntFishProducts+CntSweetProducts+CntPremiumProds_Total
avg_item_purchases	total_item_purchases/ total_purchases
Member_days_since	30/06/2024 - EnrolmentDate
dailyest_wine_purchases	CntWines/ Member_days_since

Dropped Column	
Column Name	**Justification**
CntPremiumProds	Combined into CntPremiumProds_Total
CntPremiumProds.1	Combined into CntPremiumProds_Total
Recency	Discrepancy noted when comparing against EnrolmentDate

Dropped Rows	
Column Reference	**Range Removed**
Income	24 NA count removed
EnrolmentDate	Date after 30 Jun 2024 removed
purchase_freq	Rows where purchase_freq > 0 but Total_product_purchases = 0 removed

SUMMARY STATISTICS



Product Category:

- Wines and Premium Products is the top-performing categories in terms of purchase quantity.
- Wines show the highest mean and median values (boxed in red), followed by Premium Products.
- The subsequent analysis will focus on the Wines category

	Income	Kidhome	Teenhome	CntWines	CntFruits	CntMeatProducts	CntFishProducts	CntSweetProducts	NumDealsPurchases	NumWebPurchases	NumAppPurchases
count	$ 2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091
mean	$ 52,180	0	1	311	27	169	38	27	2	4	7
std	$ 25,215	1	1	341	40	225	55	41	2	3	5
min	$ 2,447	-	-	-	-	-	-	-	-	-	-
25%	$ 35,221	-	-	25	2	16	3	1	1	2	3
50%	$ 51,315	-	-	182	8	69	12	9	2	4	6
75%	$ 68,641	1	1	516	33	236	50	35	3	6	10
max	$ 666,666	2	2	1,493	199	1,725	259	262	15	27	28

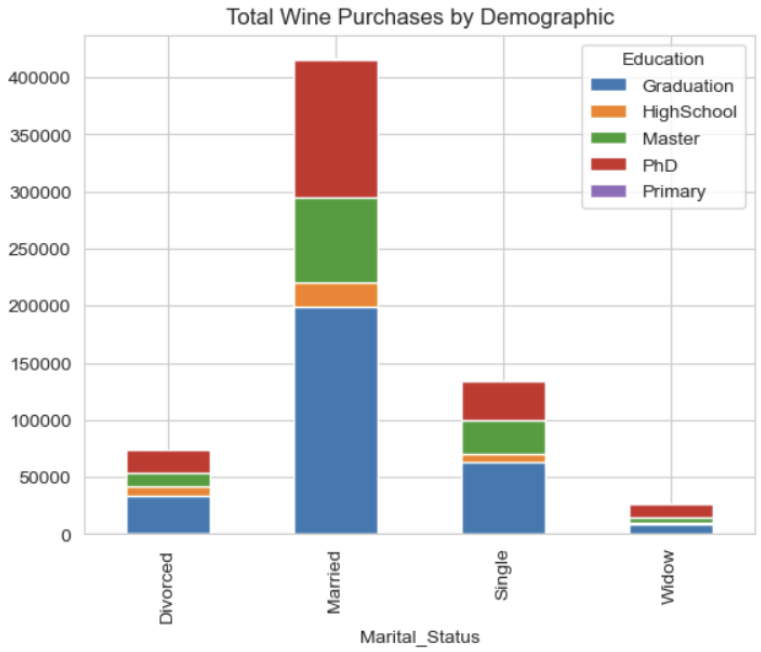
	NumStorePurchases	NumWebVisitsMonth	Complain	NumCampaign	total_kids	purchase_freq	CntPremiumProds_Total	total_product_purchases	avg_item_purchases	member_days_since	dailyest_wine_purchases
count	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091	2,091
mean	6	5	0	0	1	17	290	862	43	379	2
std	3	2	0	1	1	9	356	795	33	213	16
min	-	-	-	-	-	1	-	10	3	-	-
25%	3	3	-	-	-	8	56	147	19	211	0
50%	5	6	-	-	1	17	160	612	34	376	1
75%	8	7	-	1	1	24	369	1,436	60	537	1
max	13	20	1	5	3	54	2,619	3,760	561	905	491

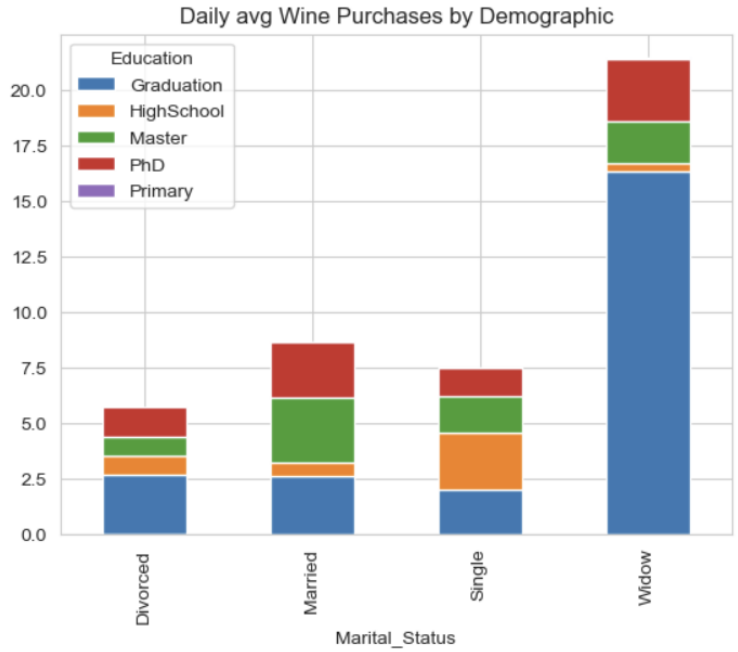
PRODUCT SALES BY CATEGORY

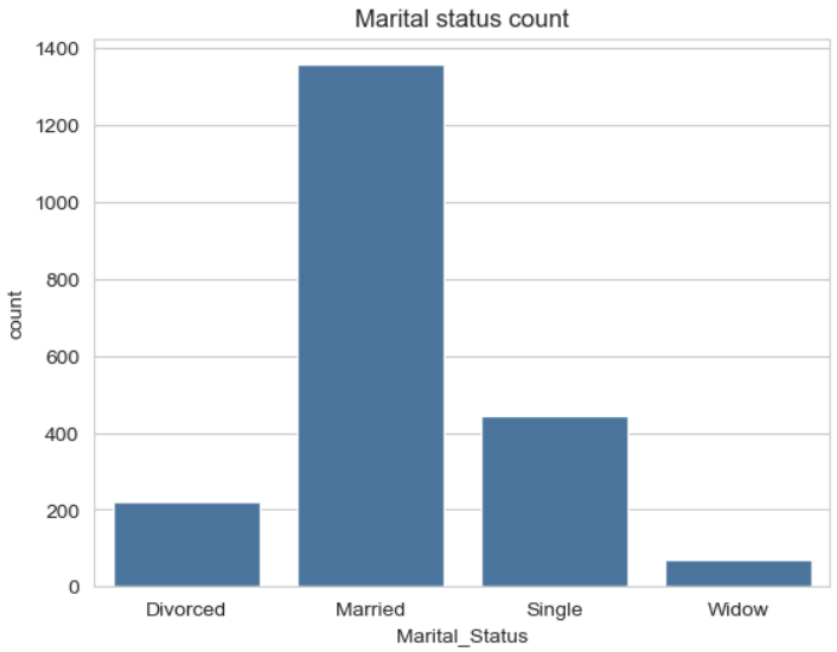
Average daily wine purchase*= (Total wine purchase count) ÷ (Days since enrolment, as of 30 June 2024).

Designed to reduce bias from purchase volumes skewed by inactive or exited members.

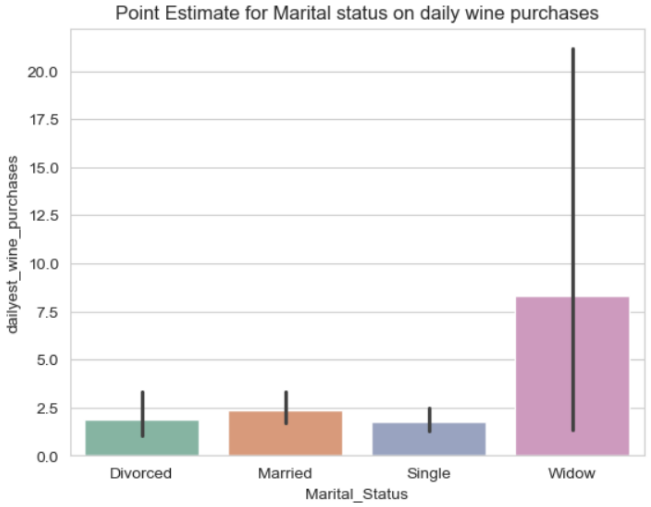
- Married customers dominate total wine sales
- Widowed customers lead in average daily purchases*.
- Though a small group (≈3% of users), widows show strong buying potential.
- Most purchases within this group come from those with a **Graduation** education level.







POINT ESTIMATES FOR WINE PURCHASE



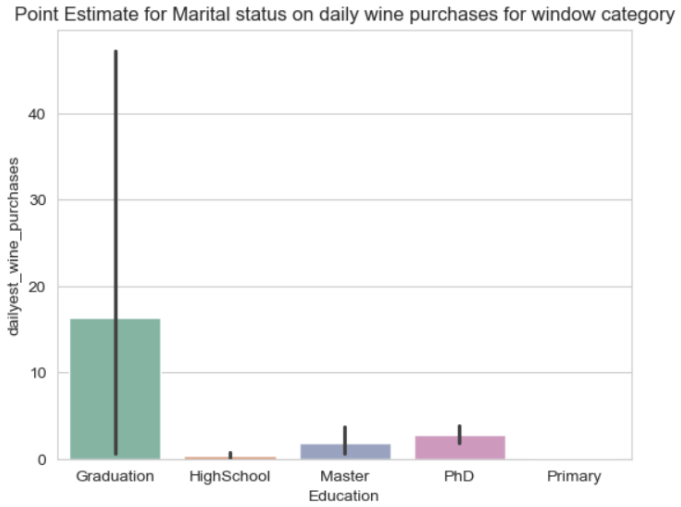


Table A: Point Estimate for Marital status on daily wine purchases

Marital_Status	sample_mean	sample_std	n	SE	CI_lower	CI_upper
Divorced	1.88	10.17	220	0.69	0.53	3.24
Married	2.39	15.61	1,357	0.42	1.56	3.22
Single	1.77	6.61	445	0.31	1.15	2.38
Widow	8.31	50.22	69	6.05	-	20.37

Table B: Point Estimate for Marital status on daily wine purchases for window category

Education	sample_mean	sample_std	n	SE	CI_lower	CI_upper
Graduation	16.30	76.09	30	13.89	-	44.71
HighSchool	0.42	0.47	5	0.21	-	1.00
Master	1.85	2.94	11	0.89	-	3.82
PhD	2.82	2.40	22	0.51	1.75	3.88
Primary	0.03		1			

<u>Observations</u>:

- The 95% confidence interval for daily wine purchases within the Widow category is [0, 20.37].
- Focusing further on widowed customers with a Graduation education level, the confidence interval widens to [0, 44.71].

<u>Issues</u>:

- Due to the relatively small sample size within the Widow category, the confidence interval remains wide and less precise.
- This estimate is expected to improve as FairFood accumulates more members under this demographic segment.
- This limitation is also reflected in the bar plot visualization and the larger standard error values observed in Table A.

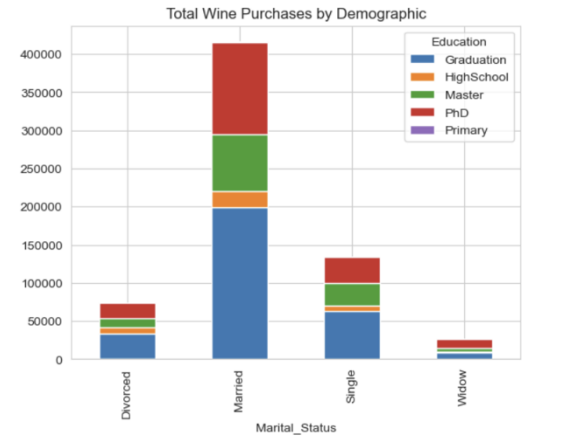
Confidence Interval with negative lower interval have been replaced with 0

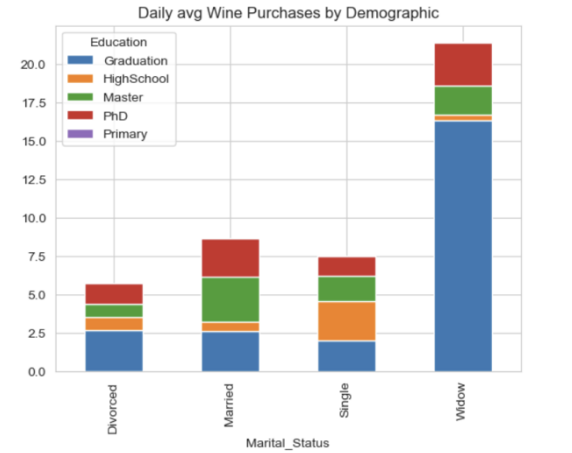
DOES "WIDOWED" PURCHASE MORE WINE THAN "MARRIED"

Normality Test: Shapiro	
Married Purchasers	
H0:	The distribution of daily wine purchases by **Married** purchasers is normally distributed.
H1:	The distribution of daily wine purchases by **Married** purchasers is <u>not</u> normally distributed.
Widowed Purchasers	
H0:	The distribution of daily wine purchases by **Widowed** purchasers is normally distributed.
H1:	The distribution of daily wine purchases by **Widowed** purchasers is <u>not</u> normally distributed.
Overall Decision:	Since pValue is **0.00 < 0.05 for both Married and Widowed**, we **reject H0** and conclude that 'dailyest_wine_purchases' distribution is not normally distributed.

Equal Variance Test: Levene	
H0:	The population variances of daily wine purchases are equal across marital groups.
H1:	The population variances of daily wine purchases are <u>not</u> equal across marital groups.
Decision:	Since pValue is **0.0122 < 0.05**, we **reject H0** and conclude that the variances 'dailyest_wine_purchases' are not equal.

Mann-Whitney U Test	
H0:	Widowed customers' wine purchases are **not** greater than married customers' purchases.
H1:	Widowed customers' wine purchases are greater than married customers' purchases.
Decision:	Since pValue is **0.0205 < 0.05**, we **reject H0** and conclude that Widowed customers' wine purchases are greater than married customers' purchases.





Hypothesis Testing Results:

- A **Shapiro–Wilk test** and a **Levene's test** were first conducted to access the normality and population variance.
- The results indicated that the **assumptions of normality and equal variances were not met**.
- Consequently, a **Mann-Whitney U Test** was performed as the non-parametric alternative.
- The resulting p-value of **0.0205** and test statistic of **53,637** suggest that the **Widowed** group purchases significantly more wine on average per day compared to the **Married** group.

SUMMARY

- In this analysis, we began by examining the overall product sales quantity across different product categories. The results showed that **Wines currently lead in total sales volume**, followed by Premium Products and Meat. This trend is consistent with the **higher mean and median** purchase quantities observed in the respective product categories.
- Focusing on the Wines category, we further analyzed **sales volume by marital status and education level**. Overall, the **Married group recorded the highest total wine purchases**. However, this is partly attributed to the larger number of users within this group, which naturally contributes to a higher cumulative total. In contrast, the **Widowed group represents only about 3%** of the total user base.
- When examining the average daily wine purchase by marital status, the **Widowed group exhibited the highest daily purchase rate**. Within this demographic, individuals with a **Graduation-level education** showed the greatest average daily purchase.
- To determine whether these observed differences were statistically significant, a **Mann-Whitney U Test** was conducted, as the data **violated the assumptions of normality and homogeneity of variance**. The test result (**p = 0.0205**) indicates the **average daily wine purchase is significantly higher among the Widowed group** compared to the Married Group**.**

Conclusion:

- The **absence of product unit price** data limits our ability to analyze total revenue and assess whether returns justify the investment costs.
- While the total wine purchases to date are considerably higher among the Married group than the Widowed group, the **daily wine purchase rate** reveals the opposite trend. The Mann–Whitney U statistic of **53,637** (p = 0.0205) indicates that widowed customers purchase significantly more wine on a daily basis compared to married customers
- This presents a potential **market opportunity** for FairFood to explore. Although the Widowed segment represents only about **3% of the total customer base**, their higher purchasing intensity suggests a valuable niche worth targeted marketing consideration.