Customer Personality Analysis

Gunay Azizova & Marc Bouché

DATA SET CHOICE

We have selected a data set from Kaggle.

Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.

- Demographic data
- Purchase channels
- Purchase amount
- Purchase habits

BUSINESS PROBLEM & HYPOTHESIS

‘We are preparing to launch a new product line, and by analyzing our existing customer database, we aim to identify their preferences and purchase habits.'

H1: There is a significant difference in purchasing habits across different customer segments(education, marital status, age), indicating that certain groups have distinct preferences.

H2: Recent customers are more likely to engage in future purchases and should be targeted in upcoming marketing campaigns.

Cleaning & Grouping the Data on SQL

```sql
delete from marketing_campaign where Education is null;
```

```sql
ALTER TABLE marketing_campaign
DROP COLUMN AcceptedCmp4;
```

```sql
ALTER TABLE marketing_campaign
rename column Amount_fruits to Fruits_products_spending;
```

```sql
UPDATE marketing_campaign
SET Education = CASE
    WHEN Education = 'Graduation' THEN 'Bachelor'
    ELSE Education
END;
```

```sql
UPDATE marketing_campaign
SET Education = CASE
    WHEN Education = '2n Cycle' THEN NULL
    ELSE Education
END;
```

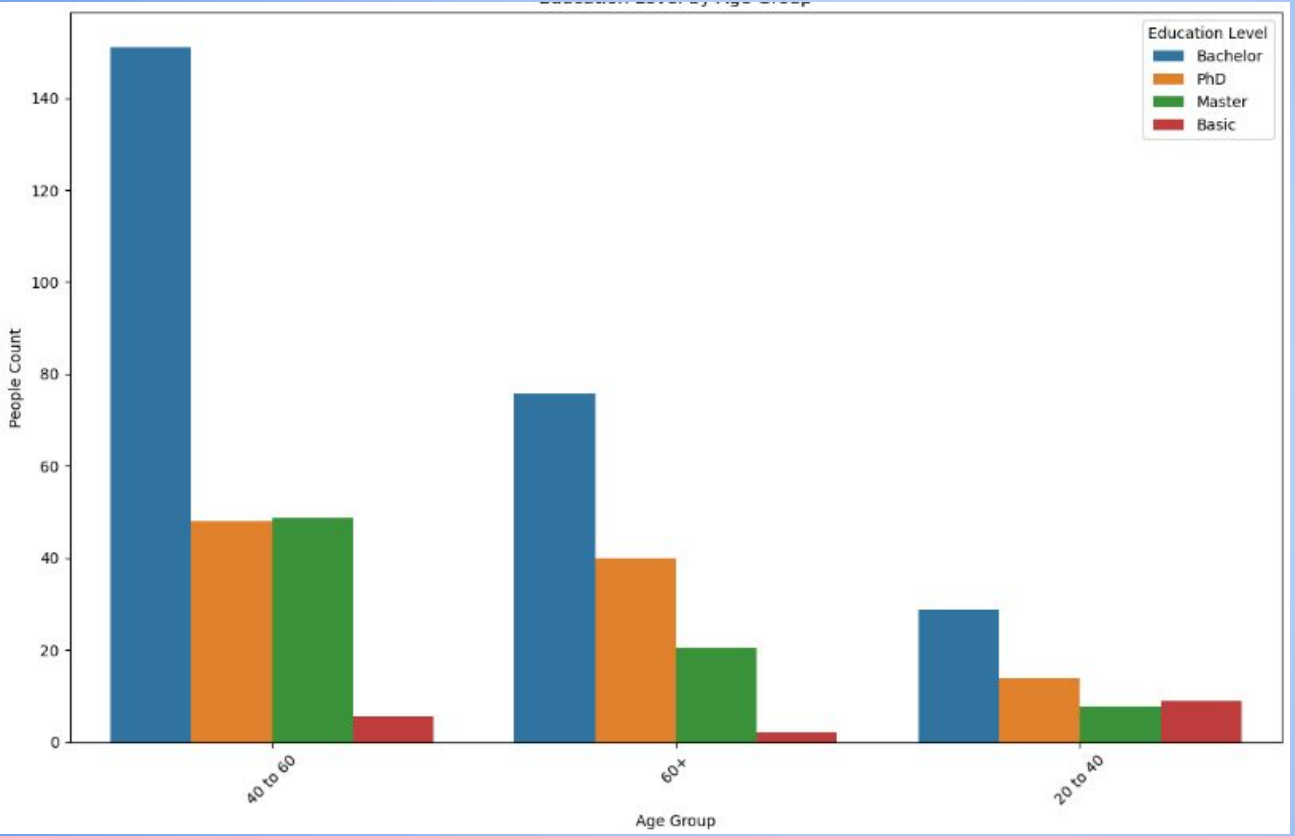
```sql
FROM (
    SELECT
        Recency,
        CASE
            WHEN Recency <= 30 THEN 'Very Recent'
            WHEN Recency BETWEEN 31 AND 60 THEN 'Recent'
            ELSE 'Older'
        END AS RecencySegment,
        Web_Purchases,
        Catalog_Purchases,
        Store_Purchases
    FROM
        marketing_campaign
) AS segmented_customers
GROUP BY
    RecencySegment;
```

H1: There is a significant difference in purchasing habits across different customer segments.

Following categories have been analysed:

- Customer Overview: Education, marital status, age (20 to 40, 40 to 60, 60+);
- Spending by Education;
- Spending by Marital status;
- Spending by Age

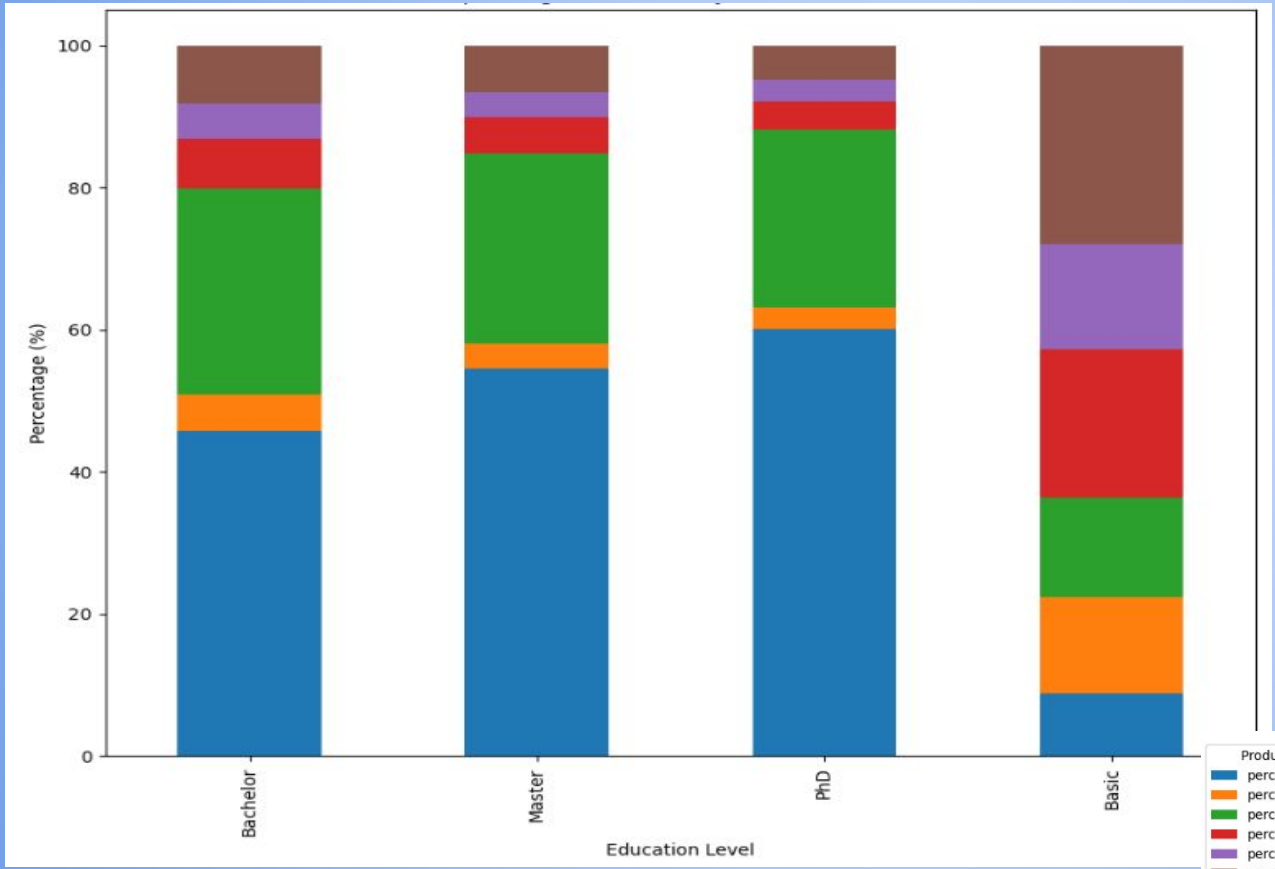
Customer Count by Education, Age, Marital Status



The largest customer group falls within the 40 to 60 age range, while the 20 to 40 age group represents the smallest segment.

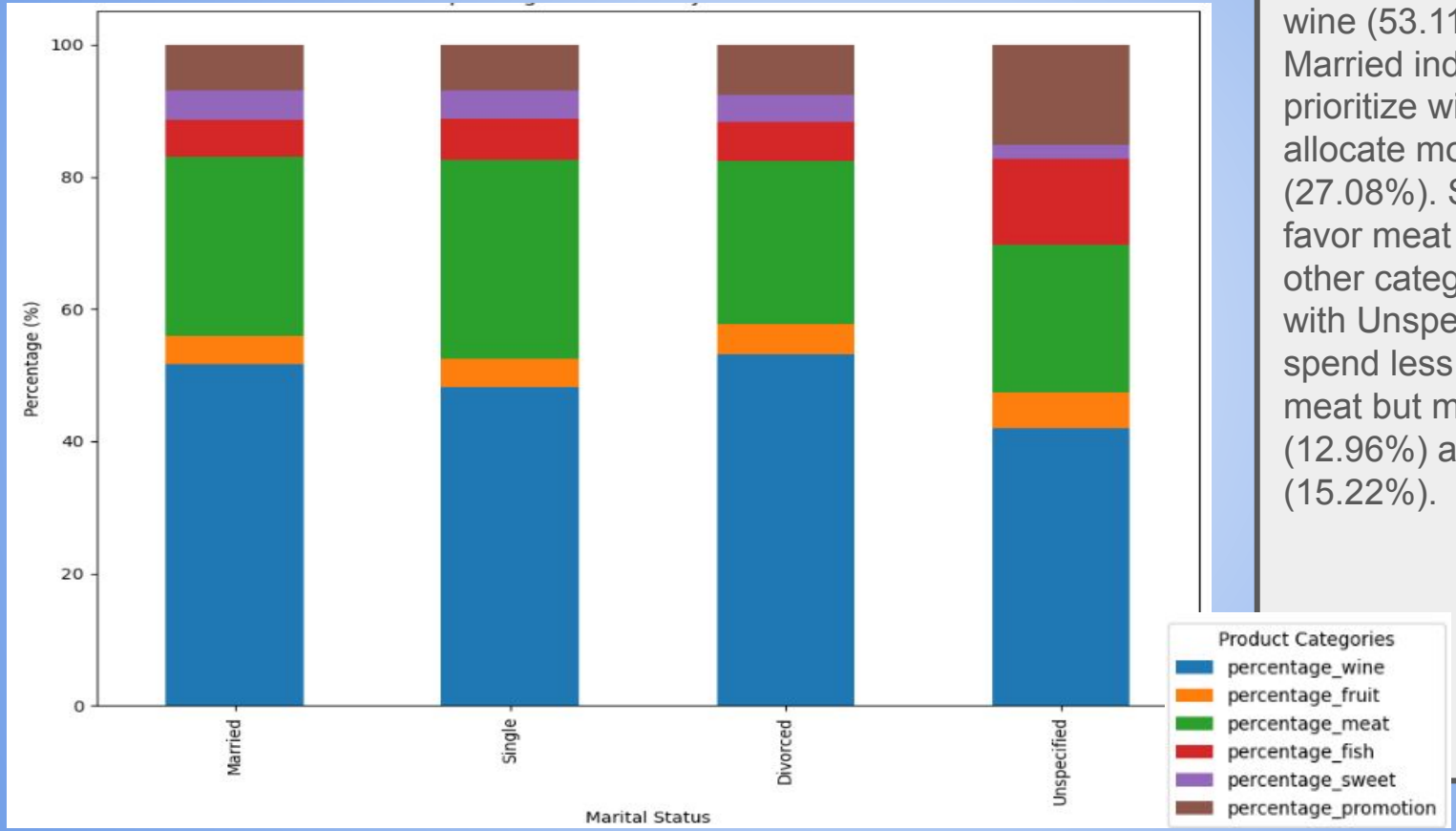
As we saw that age group customer count is significantly different, we analyze by each category separately -->

Spending Distribution by Education



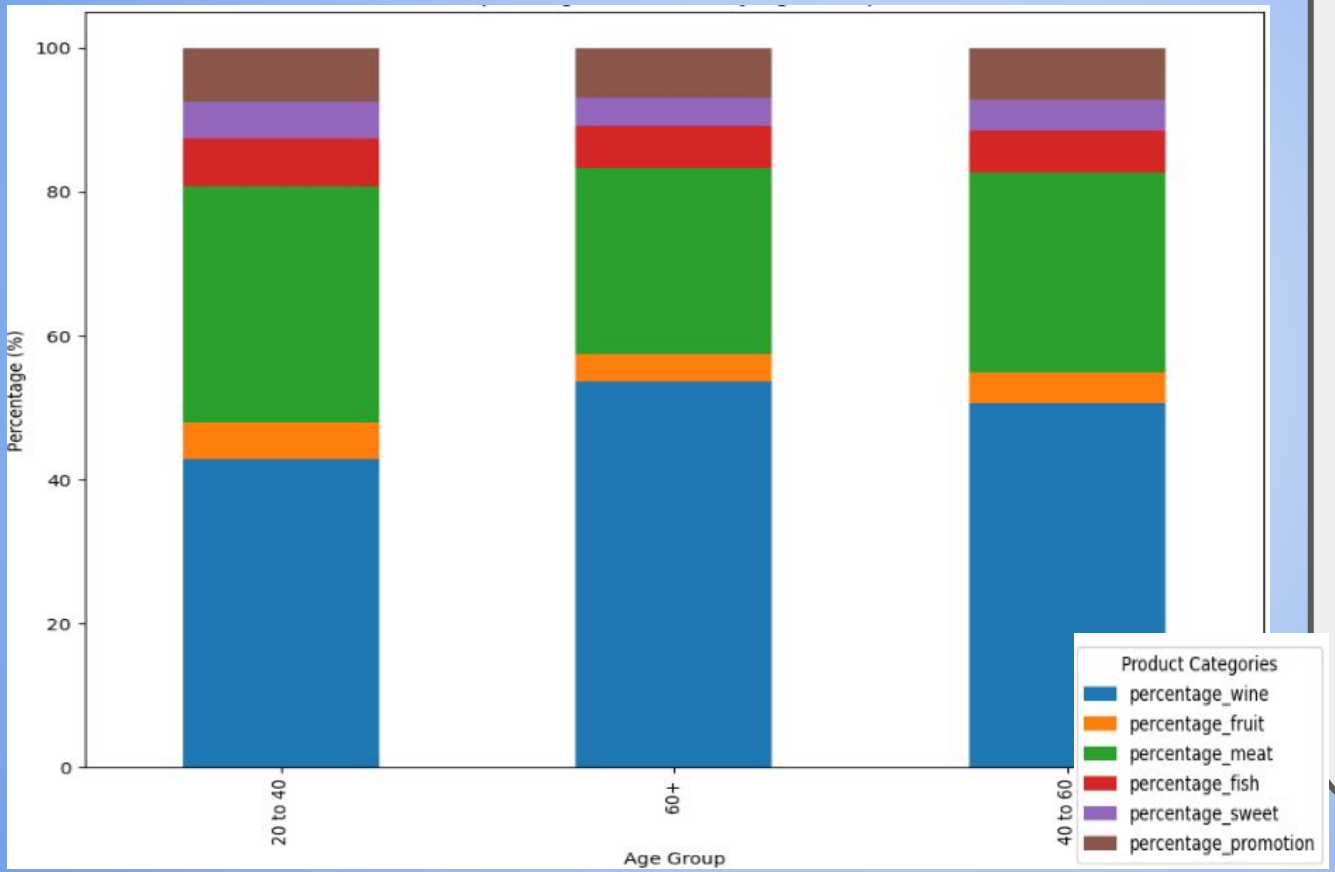
PhD holders allocate the highest percentage to wine (60.18%), while those with Basic education spend significantly more on fruits (13.58%), fish (20.85%), and promotions (27.91%). In contrast, Bachelor's and Master's degree holders fall between these extremes, with a higher focus on wine and meat but less on other categories.

Spending Distribution by Marital Status



Divorced customers spend the highest percentage on wine (53.11%), while Married individuals also prioritize wine (51.72%) but allocate more to meat (27.08%). Single customers favor meat (29.97%) over other categories, and those with Unspecified status spend less on wine and meat but more on fish (12.96%) and promotions (15.22%).

Spending Distribution by Age Group



The 60+ group spends the most on wine (53.74%), followed closely by the 40 to 60 group (50.64%), while the 20 to 40 group focuses more on meat (32.90%). Spending on fruits, fish, and promotions remains fairly consistent across age categories, but younger customers (20 to 40) allocate slightly more to promotions (7.51%). Knowing that the highest customer count is in 40 to 60, we observe the same purchasing preferences.

H1: There is a significant difference in purchasing habits across different customer segments. ✗

- Basic customer segment showcased different purchasing habits compared with the rest of customers by education;
- While specific segments exhibit distinct preferences in certain categories, the overarching trends in wine, meat, and other product categories suggest that purchasing habits are not significantly different across customer segments;
- The hypothesis is not strongly supported by the data except one case. Therefore, it could be partially rejected.

H2: Recent customers are more likely to engage in future purchases and should be targeted in upcoming marketing campaigns.

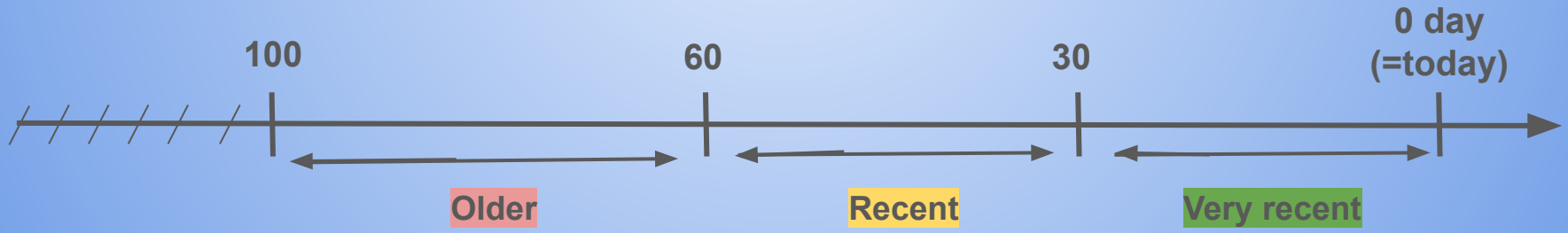
Important definition and information before visualization and results

1.The data set did not provide data for customer having a regency>100 days.

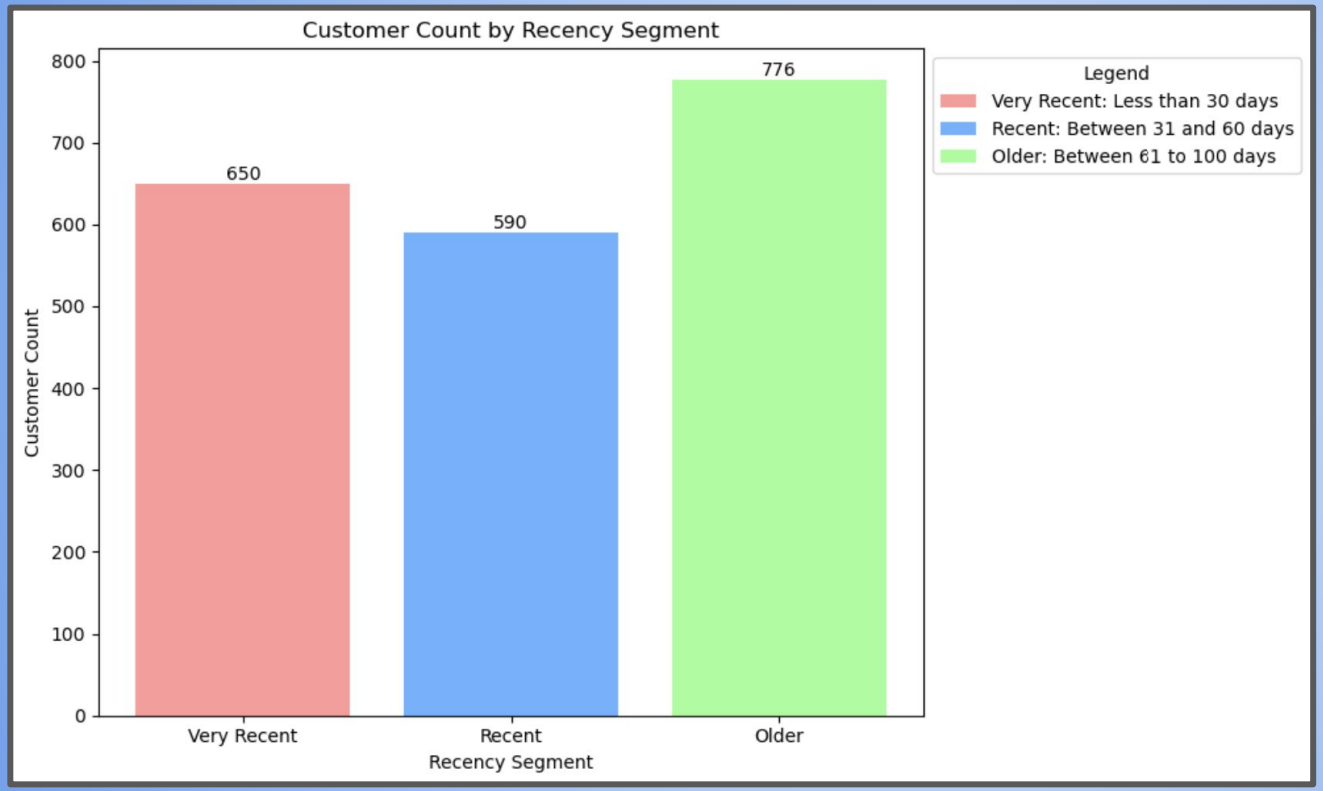
2. **Recency:** Number of days since customer's last purchase.

Recency is divided into 3 categories:

- <=30 days AS ‘Very Recent’
- 31 to 60 days AS ‘Recent’
- 61 to 100 days AS ‘Older’



Customer Count by Recency Segment

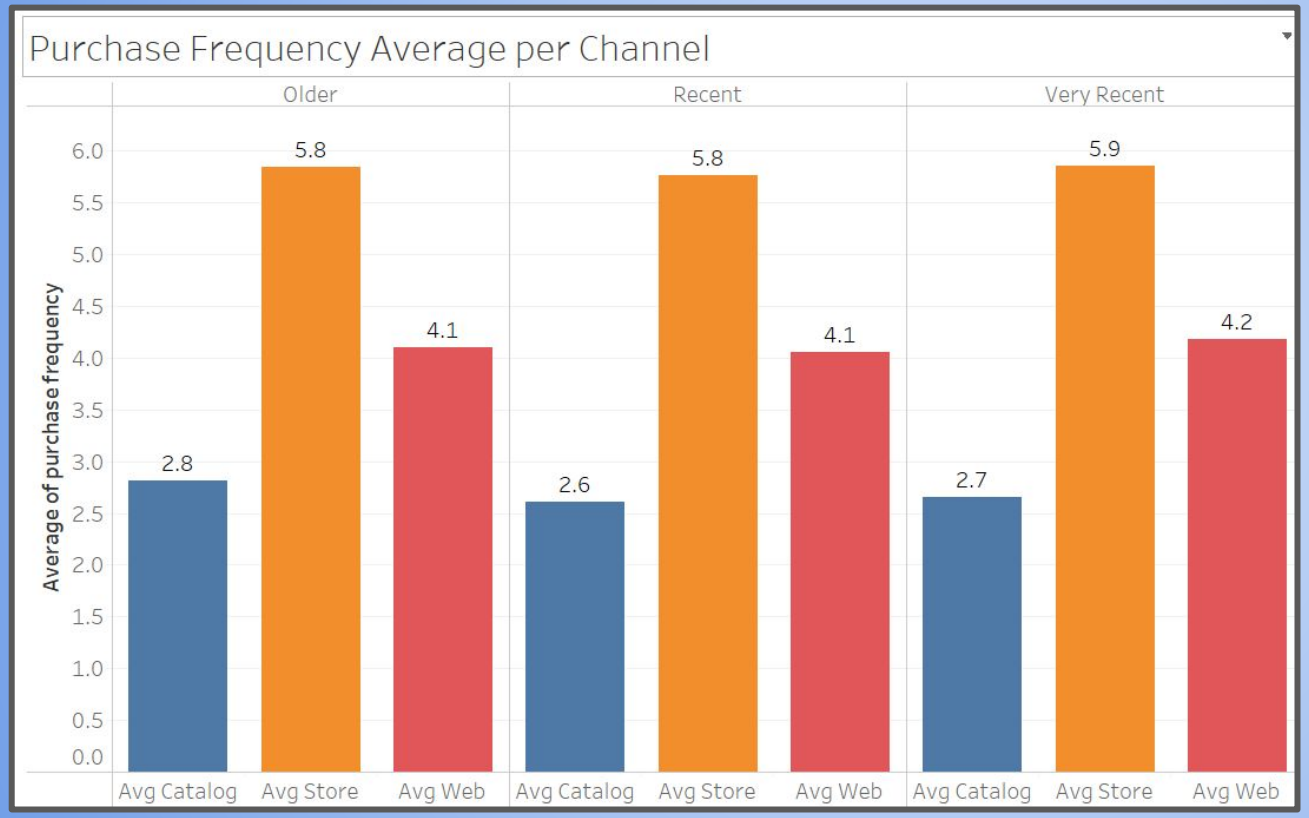


The older recency segment has the largest amount of customer with 776 customers that did not purchase anything for the past 2 months minimum.

590 customers did not purchase a thing for the past month only.

650 customers purchased articles this month.

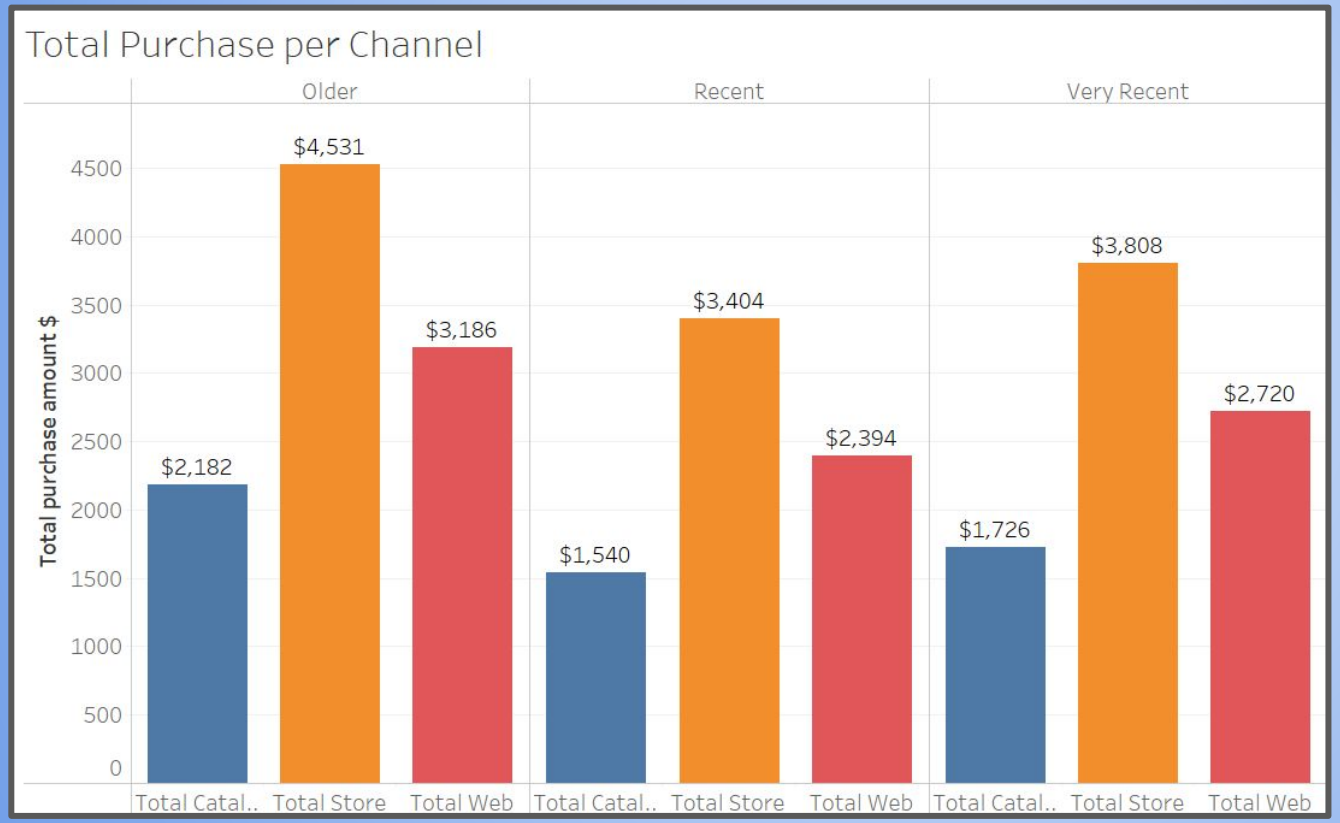
Purchase Frequency Average per Channel



There is no significant difference between each channel regarding the purchase frequency channel.

No matter how long the customer last purchased an article, we observe a real consistency in term of Channel distribution.

Total Purchase per Channel

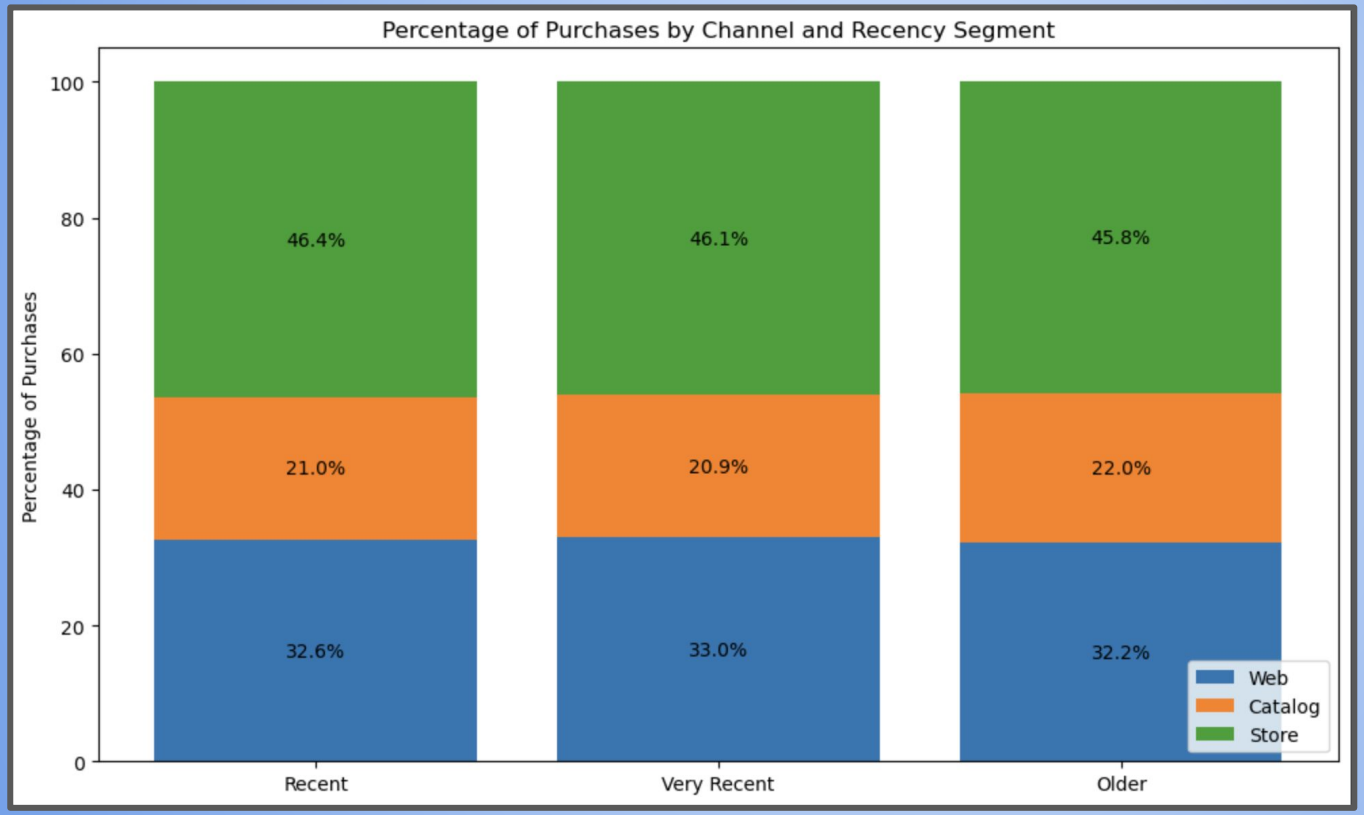


Customers tend to purchase more in stores than other channel, no matter what regency they belong to.

Catalog is the least preferred channel by the customers.

We also observing that customer from this month have spent more money across all channels.

Percentage of Purchases by Channel and Recency Segment



In-store purchases account for nearly half of all sales.

Catalog purchases, though the least popular, still represent 21% of recent customer orders.

We found no significant differences in purchasing behavior based on customer recency.

H2: Recent customers are more likely to engage in future purchases and should be targeted in upcoming marketing campaigns. ✗

- The data shows an even distribution of customers across recency segments, with no significant concentration in the "most recent" group.

- Purchase frequency and total purchases are similar across segments, and the share of purchases is evenly spread, indicating no clear advantage for targeting recent customers.

- The hypothesis is not strongly supported by the data. Therefore, it is rejected.

— CONCLUSION —

- The evidences shown by the analysis for hypothesis 1 and 2 are too weak to support it as no significative difference has been observed between neither customer segments by education, age, and marital status, as well as most recent customers and the other groups.
- Given the absence of major outliers in our customer data, retargeting efforts for the new product line can be confidently driven by existing preferences.

Major Obstacle

- GitHub

Thank you

Gunay Azizova & Marc Bouché