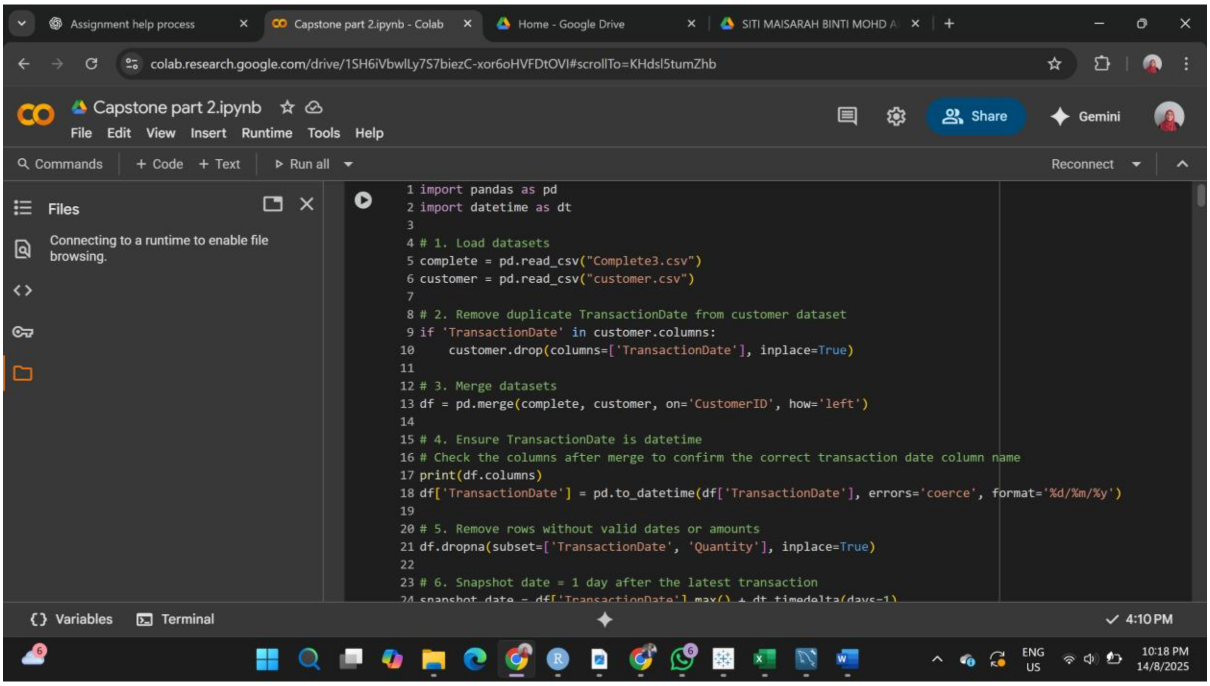
Customer Retention and Sales Optimization in Retail

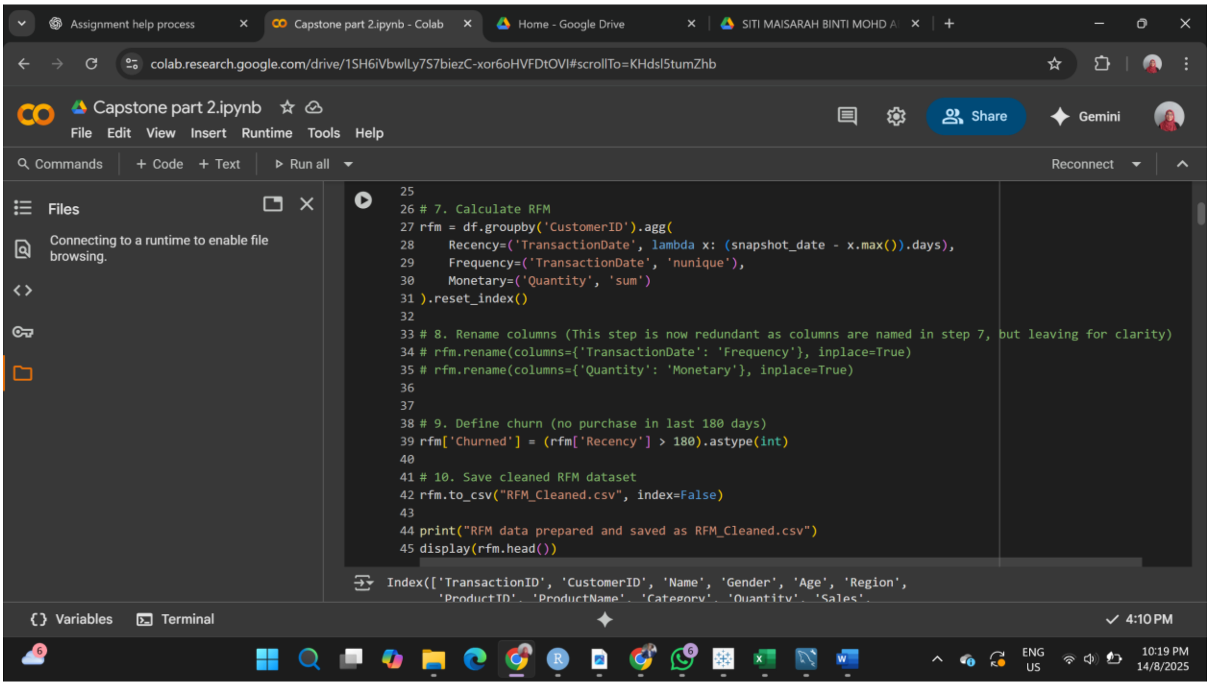
Part 2 - BI Dashboard, Data Science, & R Programming

1. Data cleaning and transformation

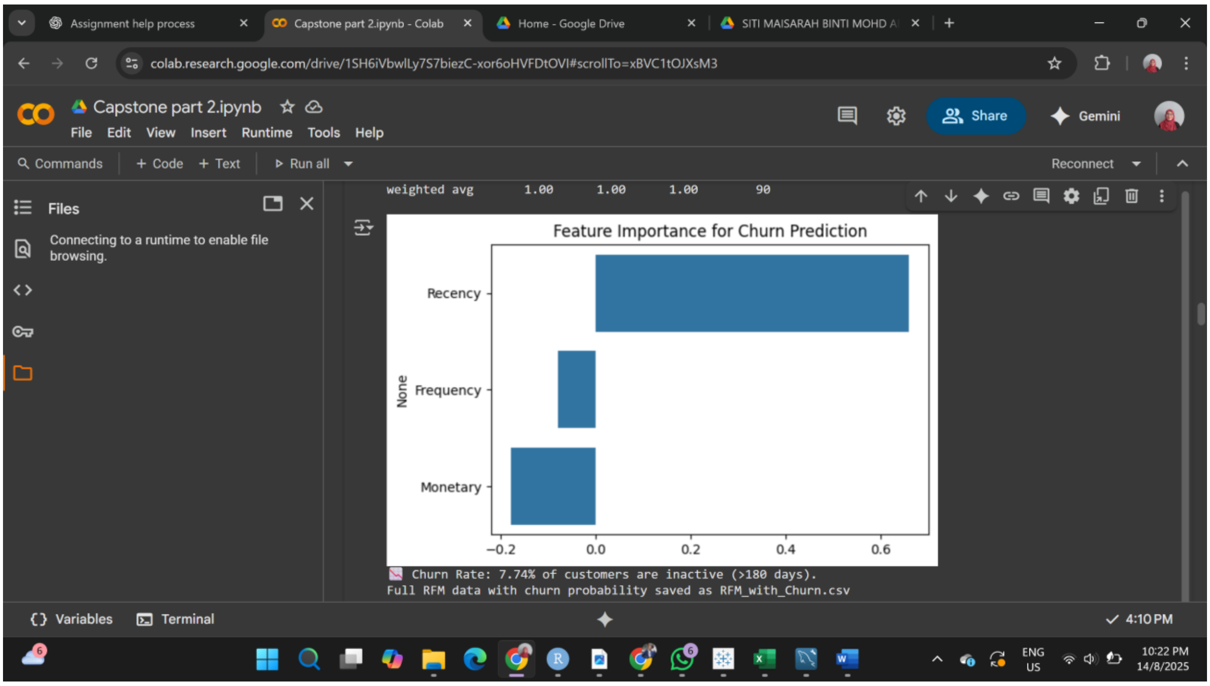
   a) Data cleaning



   b) Calculate RFM; Recency, Frequency, and Monetary



c) Feature importance for churn prediction visualisation using barplot



2. Data science – R language

a) Chi-squared test result

```
        Pearson's Chi-squared test

data:  table_data
X-squared = 0.59575, df = 5, p-value = 0.9882
```

The Chi-squared test result shows that there is no significant differences between male and female in preferred product category as the p-value (0.99) higher than 0.05.

b) ANOVA test result

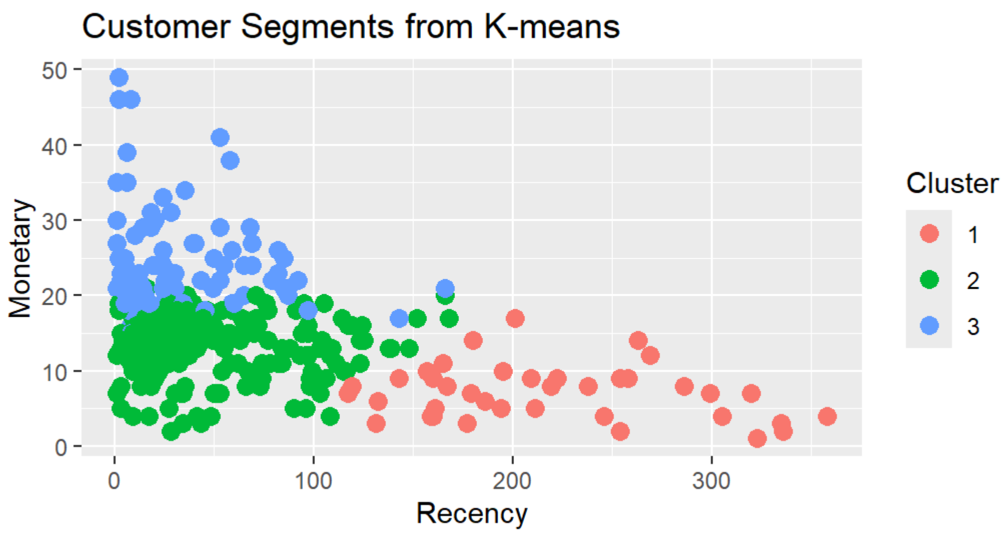
```
> anova_result <- aov(Monetary ~ Region, data = merged_data)
> summary(anova_result)
              Df Sum Sq Mean Sq F value   Pr(>F)    
Region         3   1754   584.7   8.092 2.39e-05 ***
Residuals   1496 108082    72.2                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> TukeyHSD(anova_result)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Monetary ~ Region, data = merged_data)

$Region
                  diff        lwr        upr     p adj
North-East  -0.2805521 -1.9117210  1.3506167 0.9710824
South-East  -2.8835016 -4.5273787 -1.2396245 0.0000411
West-East   -1.4058425 -2.9026698  0.0909849 0.0745775
South-North -2.6029494 -4.3273432 -0.8785556 0.0006241
West-North  -1.1252903 -2.7101228  0.4595422 0.2613520
West-South   1.4776591 -0.1202502  3.0755685 0.0816908
```

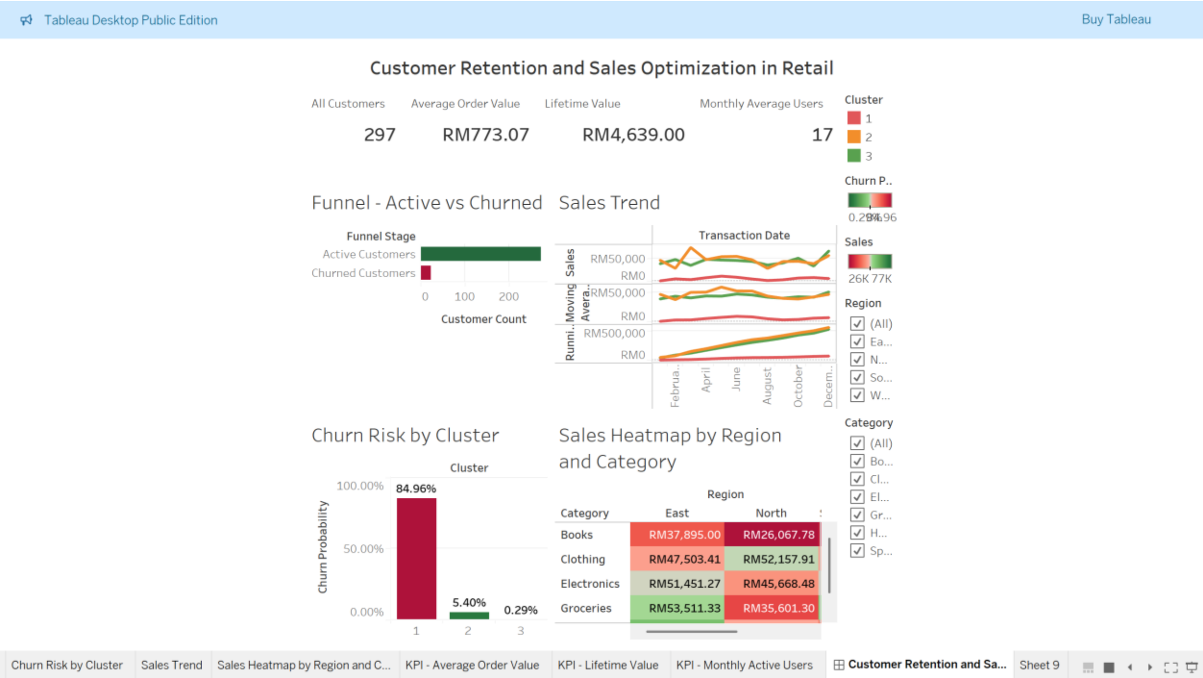
For the ANOVA result, it shows a very high significant differences for monetary value across the regions. TukeyHSD gives more in-depth result for the differences between two regions. It is found that South is significantly lower than East and North. Other that that, there is no significant difference.

c) Clustering using K-means



For cluster 1 with red color, it indicates high recency and low monetary. Most likely churned customers. For cluster 2 with green color shows moderately on both recency and monetary. For the blue one, cluster 3 is the loyal and high-value customers.

3. Visualisation using Tableau



The dashboard shows two charts focusing on churn and two charts on sales. It is also equipped by total customers, average order value, lifetime value, and monthly average users as its KPI. Region and product category are used as filters.