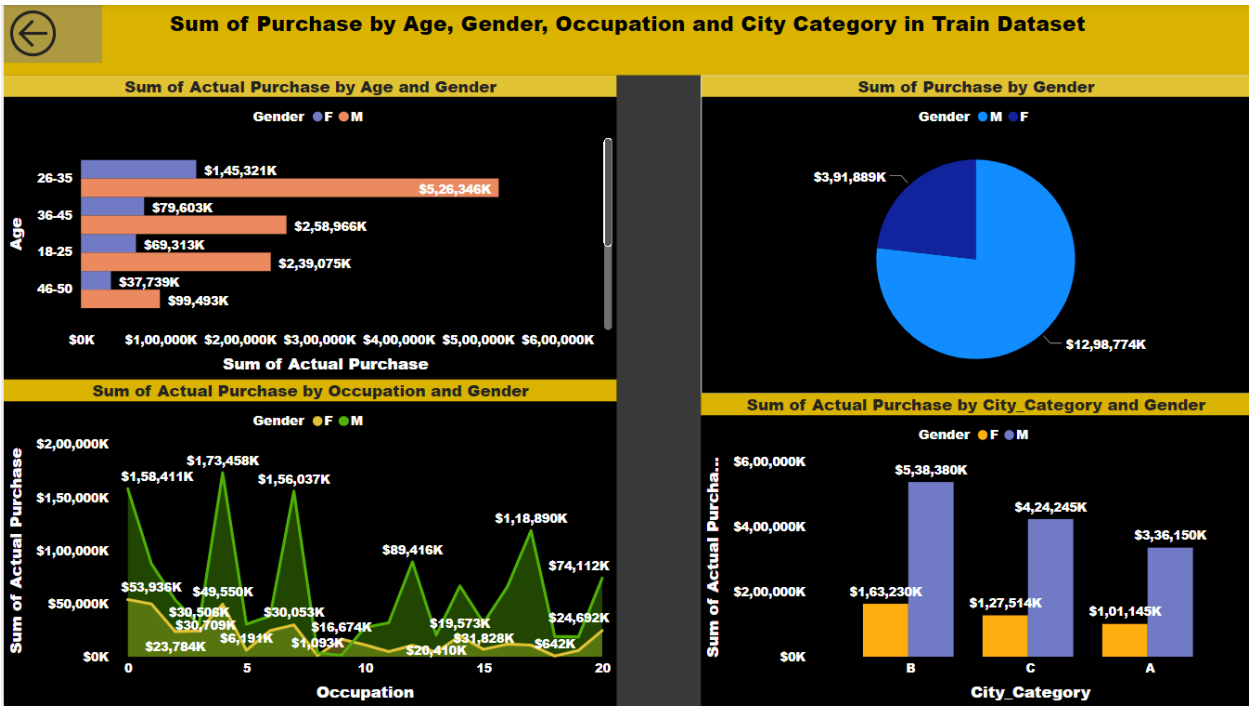
Black Friday Sales Prediction

Wireframe Documentation

Data Visualization using Power BI

During Exploratory Data Analysis, Following insights about the Train Data and Test Data has been found which are given below:

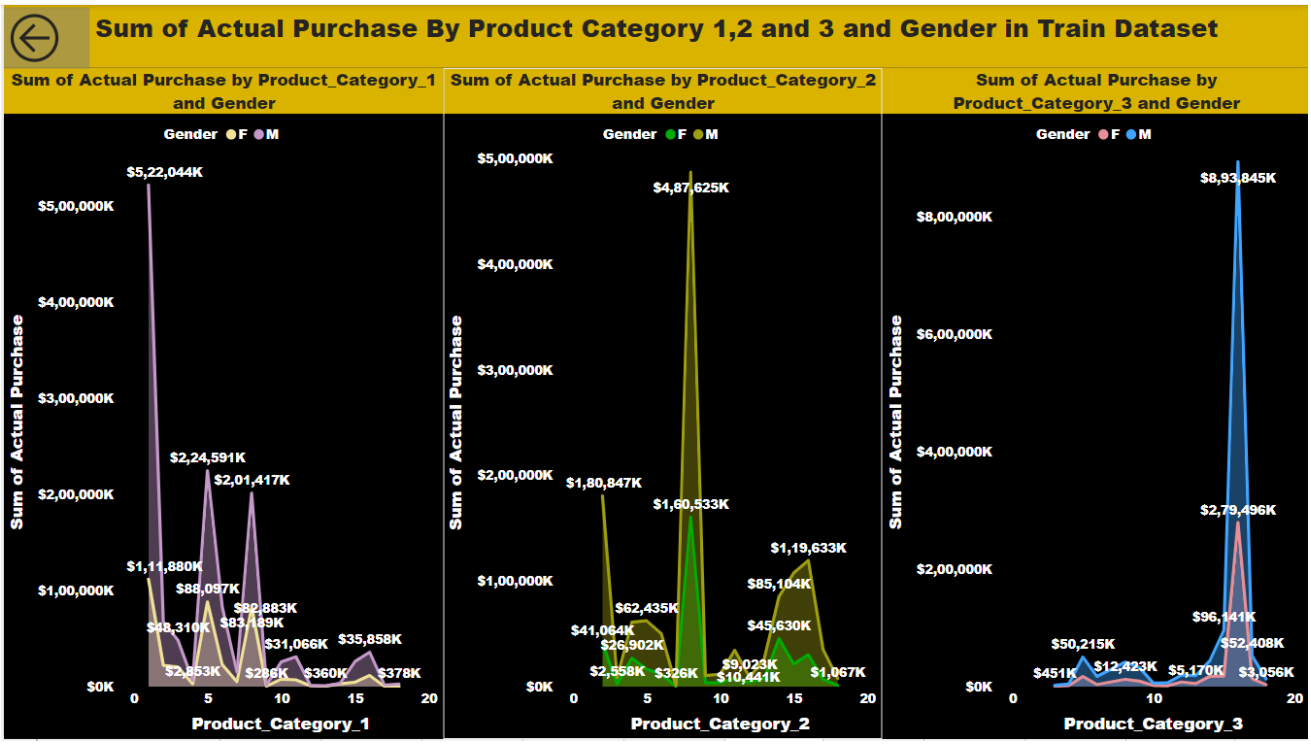
(1) Sum of Actual Purchase by Age, Gender, Occupation and City Category in Train Dataset



From the Above Insight, following data can be seen:

(a) Sum of actual purchase is highest in age group between 26-35 at $5,26,346k for Male and $1,45,321k for Female.
(b) Sum of actual purchase is $12,98,774k for male and $3,91,889k for female.
(c) Sum of actual purchase is highest in occupation 5 at $1,73,458k for male and $53,936k for female at 0 occupation.
(d) Sum of actual purchase is highest in City Category B at $5,38,380k for Male and $1,63,230k for female.

(2) Sum of Actual Purchase by Product Category 1,2 &3 in Train Dataset



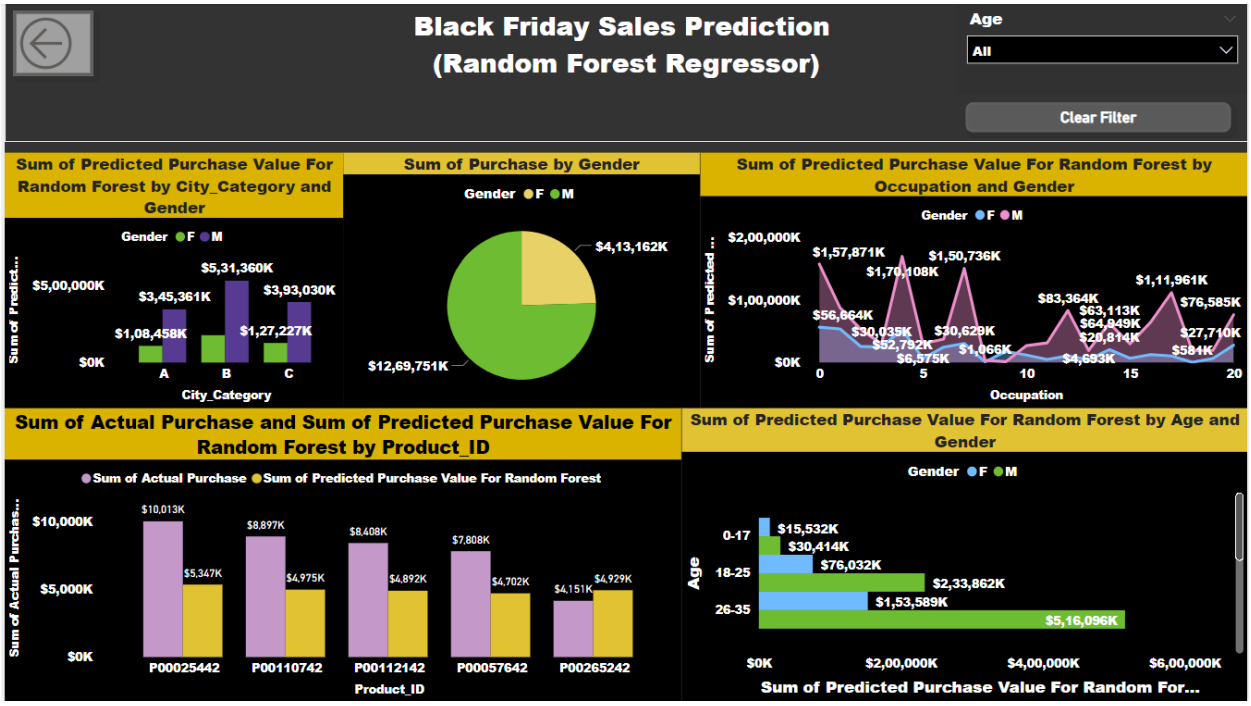
From the Above Insight, following data can be seen:

(a) Sum of Actual Purchase is highest in Product category 1 at $5,22,044k for Male in 1 and $2,24,591k for female in 5.

(b) Sum of Actual Purchase is highest in Product category 2 at $4,87,625k for Male in 8 and $1,80,847k for female in 2.

(c) Sum of Actual Purchase is highest in Product category 3 at $8,93,845k for Male in 16 and $2,79,496k for female in 16.

(3) Dashboard for Random Forest Regressor



From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase value for Random Forest is highest with $5,31,360k for Male and $1,77,477k in city category B .

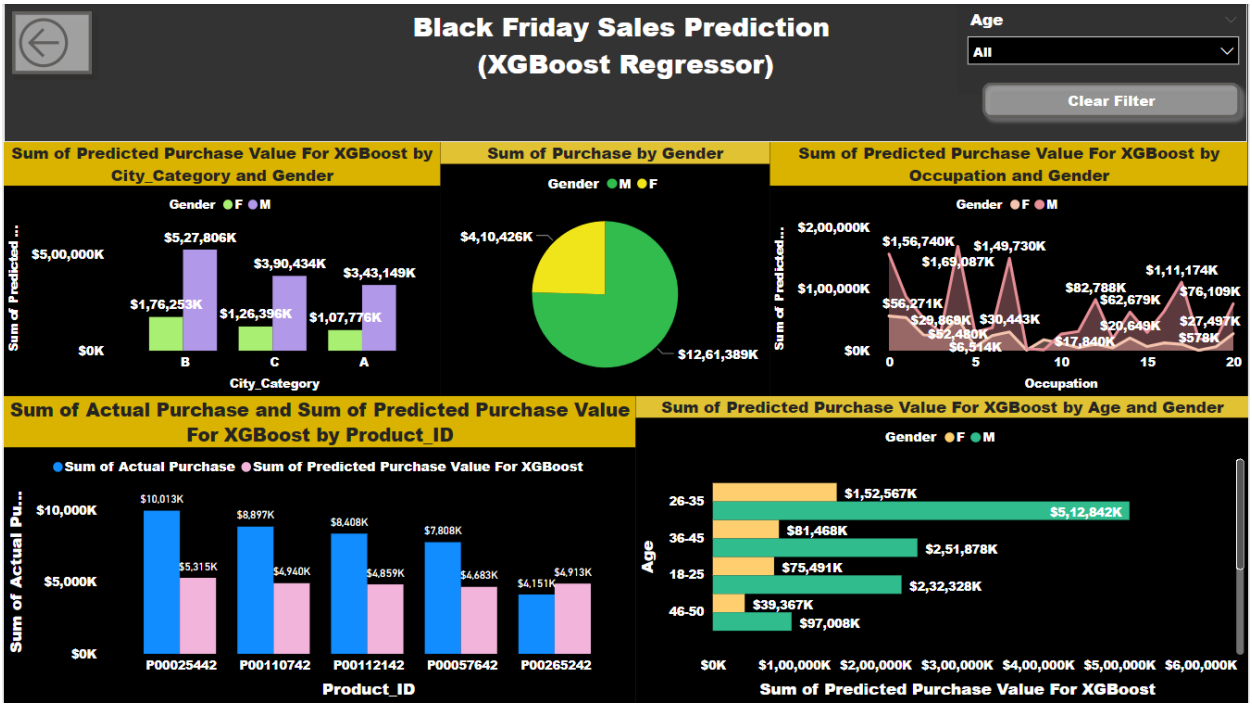
(b) Sum of purchase by gender for Random forest is $12,69,751K for Male and $4,13,162k for female.

(c) Sum of Predicted Purchase value for Random forest is highest with $1,57,871k for Male and $56,664k for female in occupation category 0.

(d) Sum of actual purchase is highest with $10,013k and Sum of predicted purchase value is highest with $5,347k for Random forest in product id P00025442.

(e) Sum of Predicted purchase value for Random Forest is highest with $5,16,096k for Male and $1,53,589k for female in the age group 26-35.

(4) Dashboard for XGBoost Regressor



From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase value for Random Forest is highest with $5,27,806k for Male and $1,76,253k in city category B .

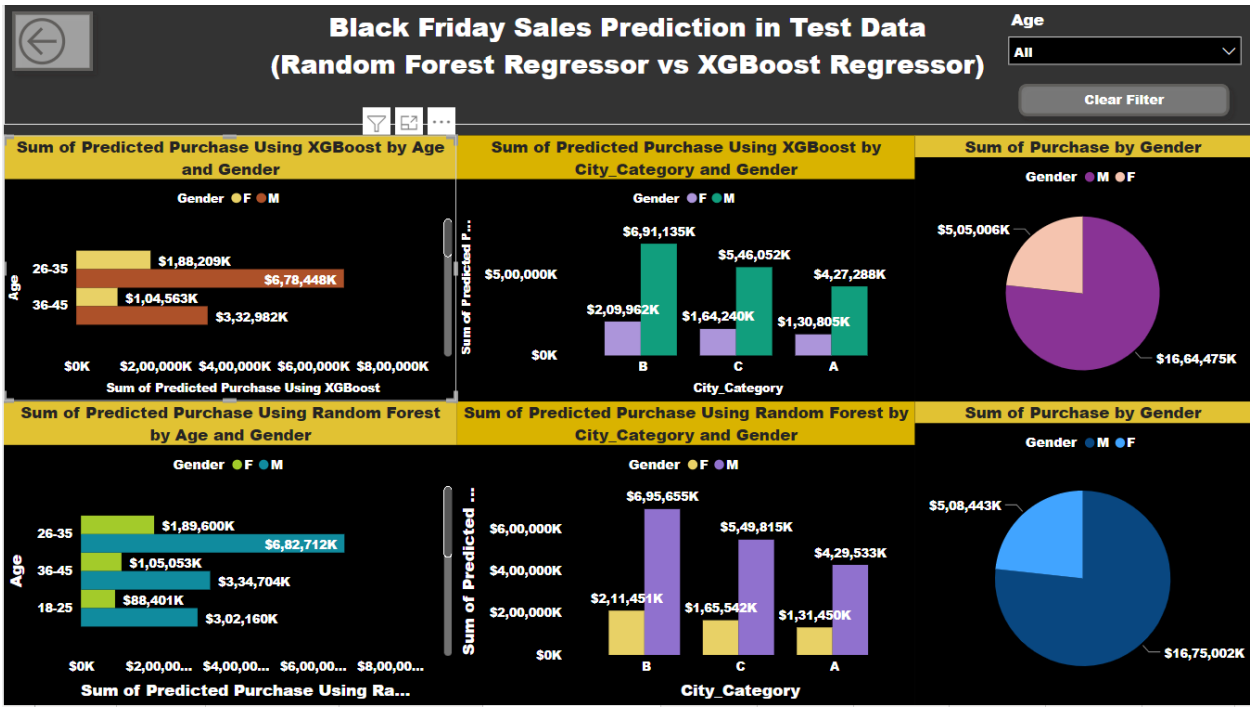
(b) Sum of purchase by gender for Random forest is $12,61,389K for Male and $4,10,426k for female.

(c) Sum of Predicted Purchase value for Random forest is highest with $1,56,740k for Male and $56,271k for female in occupation category 0.

(d) Sum of actual purchase is highest with $10,013k and Sum of predicted purchase value is highest with $5,315k for Random forest in product id P00025442.

(e) Sum of Predicted purchase value for Random Forest is highest with $5,12,842k for Male and $1,52,567k for female in the age group 26-35.

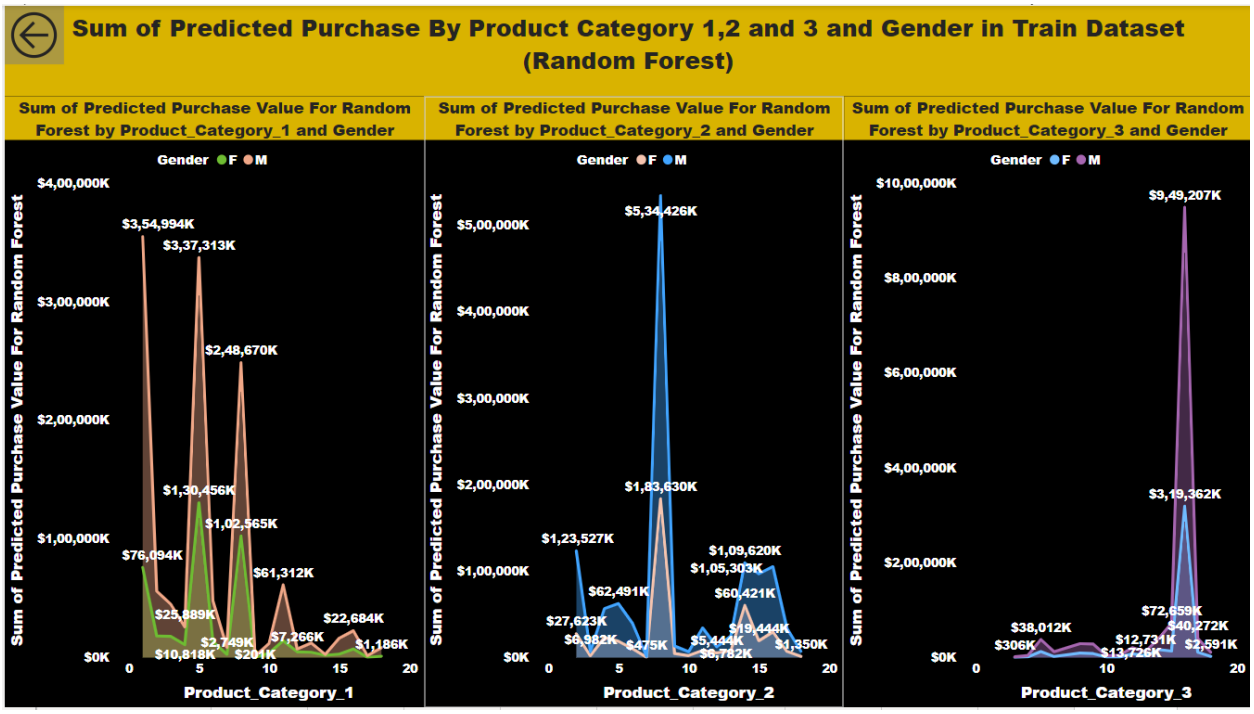
(5) **Dashboard for Test Dataset**



From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase using XGBoost regressor for test dataset is highest in the age group 26-35 at $6,78,448k for Male and $1,88,209k for Female.
(b) Sum of predicted purchase using XGBoost regressor for test dataset is highest in the city category B at $6,91,135k for Male and $2,09,962k for female.
(c) Sum of predicted purchase using XGBoost regressor for test dataset is $16,64,475k for Male and $5,05,006k for female.
(d) Sum of Predicted Purchase using Random forest regressor for test dataset is highest in the age group 26-35 at $6,82,712k for Male and $1,89,600k for Female.
(e) Sum of predicted purchase using Random forest regressor for test dataset is highest in the city category B at $6,95,655k for Male and $2,11,451k for female.
(f) Sum of predicted purchase using Random forest regressor for test dataset is $16,75,002k for Male and $5,08,443k for female.

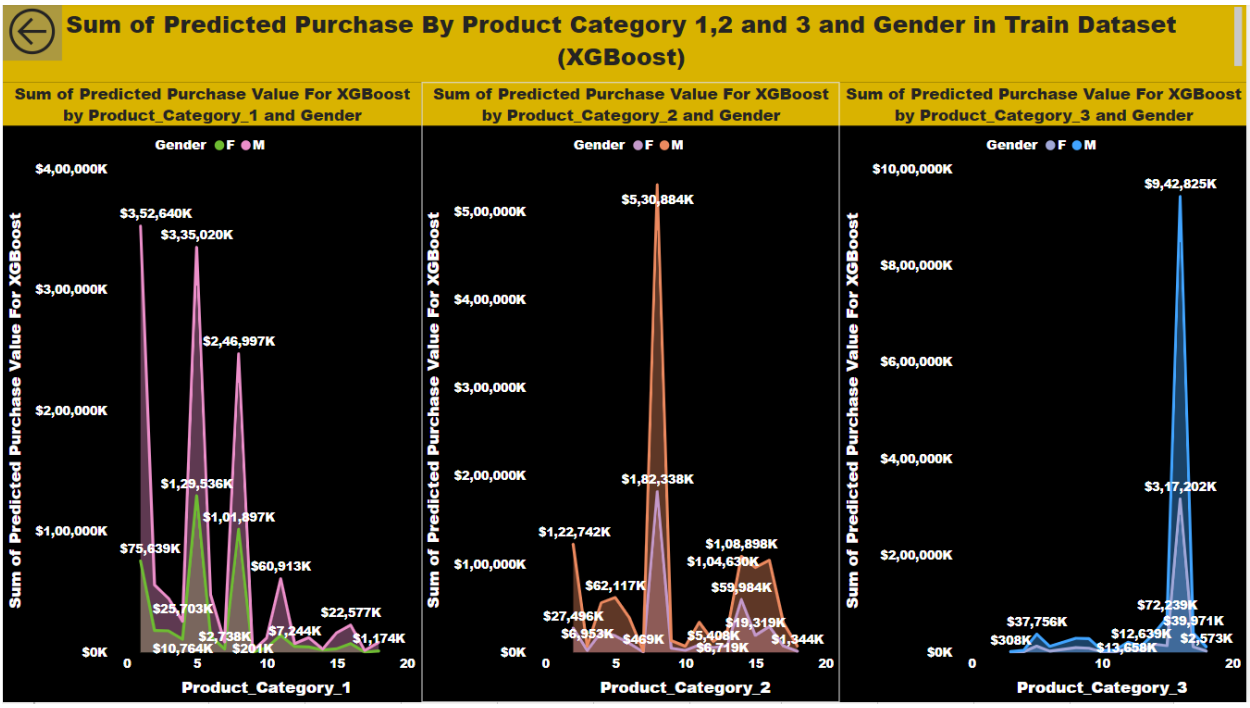
(6) Sum of Predicted Purchase by Product Category 1,2 and 3 in Train Dataset for Random Forest



From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase is highest in Product category 1 at $3,54,994k for Male in 1 and $1,30,456k for female in 5.
(b) Sum of Predicted Purchase is highest in Product category 2 at $5,34,426k for Male in 8 and $1,83,630k for female in 8.
(c) Sum of Predicted Purchase is highest in Product category 3 at $9,49,207k for Male in 16 and $3,19,362k for female in 16.

(7) Sum of Predicted Purchase by Product Category 1,2 and 3 in Train Dataset for XGBoost



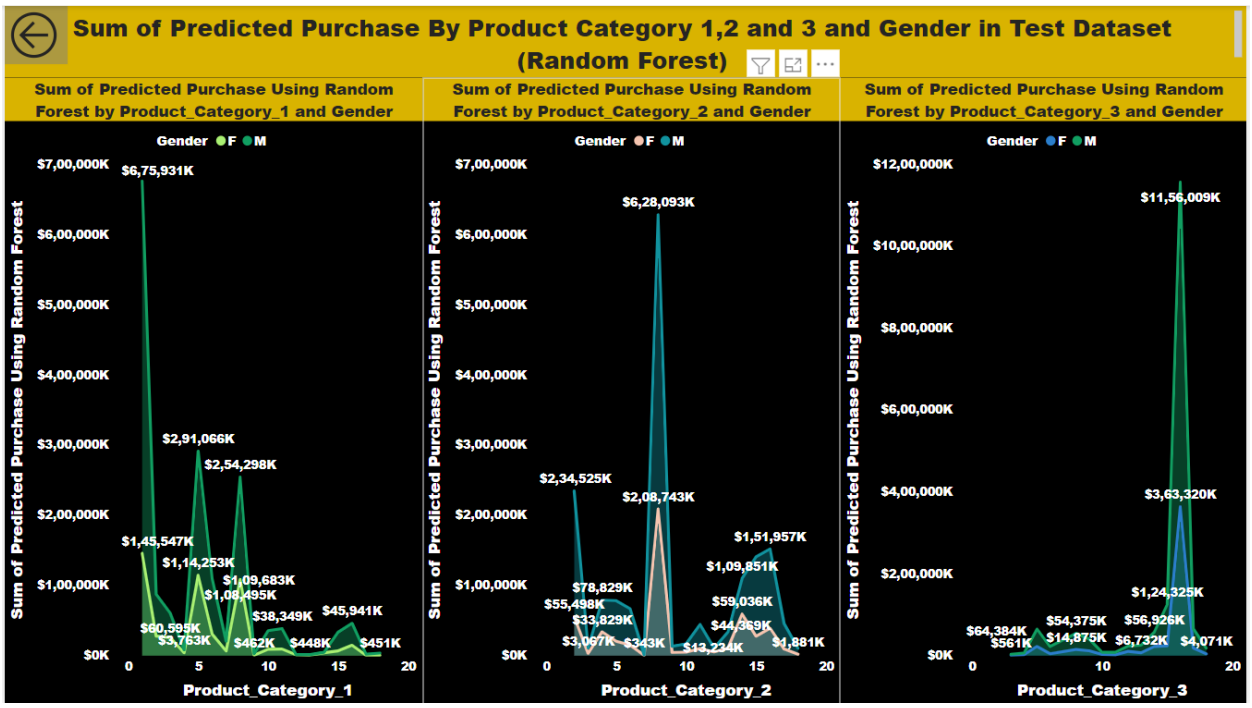
From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase is highest in Product category 1 at $3,52,640k for Male in 1 and $1,29,536k for female in 5.

(b) Sum of Predicted Purchase is highest in Product category 2 at $5,30,884k for Male in 8 and $1,82,338k for female in 8.

(c) Sum of Predicted Purchase is highest in Product category 3 at $9,42,825k for Male in 16 and $3,17,202k for female in 16.

(8) <u>Sum of Predicted Purchase by Product Category 1,2 and 3 in Test Dataset for Random forest</u>



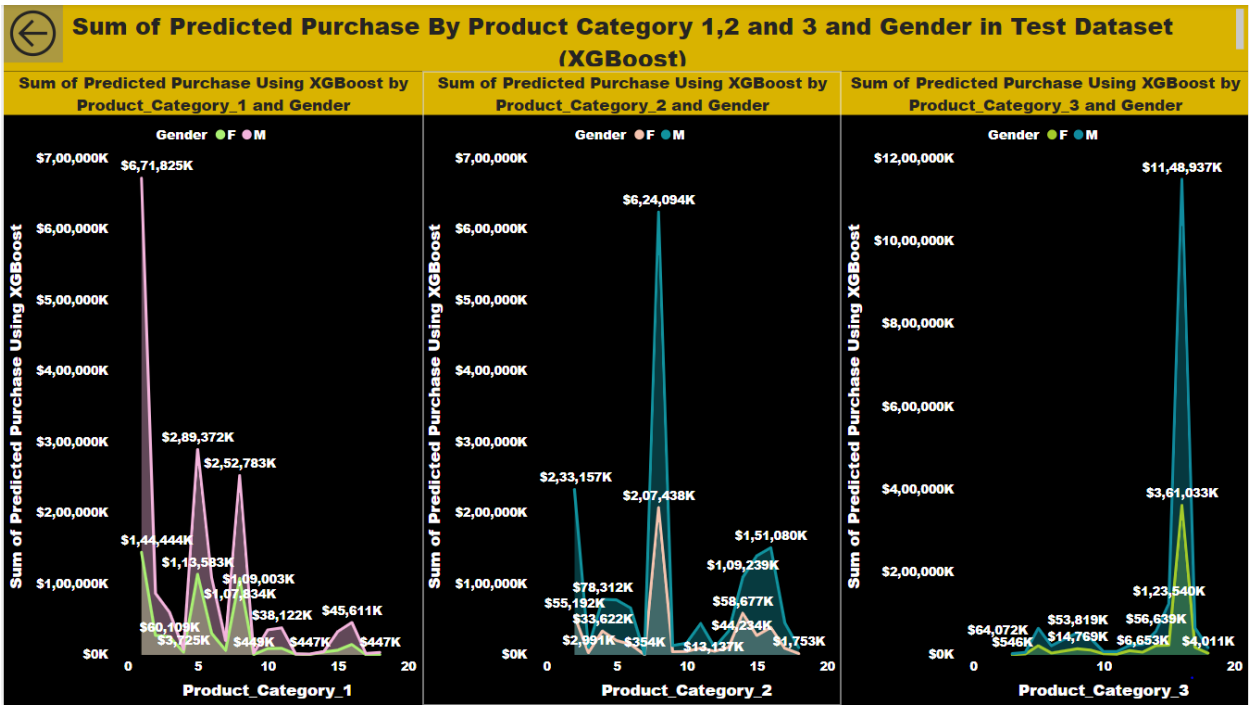
From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase is highest in Product category 1 at $6,75,931k for Male in 1 and $1,45,547k for female in 1.

(b) Sum of Predicted Purchase is highest in Product category 2 at $6,28,093k for Male in 8 and $2,08,743k for female in 8.

(c) Sum of Predicted Purchase is highest in Product category 3 at $11,56,009k for Male in 16 and $3,63,320k for female in 16.

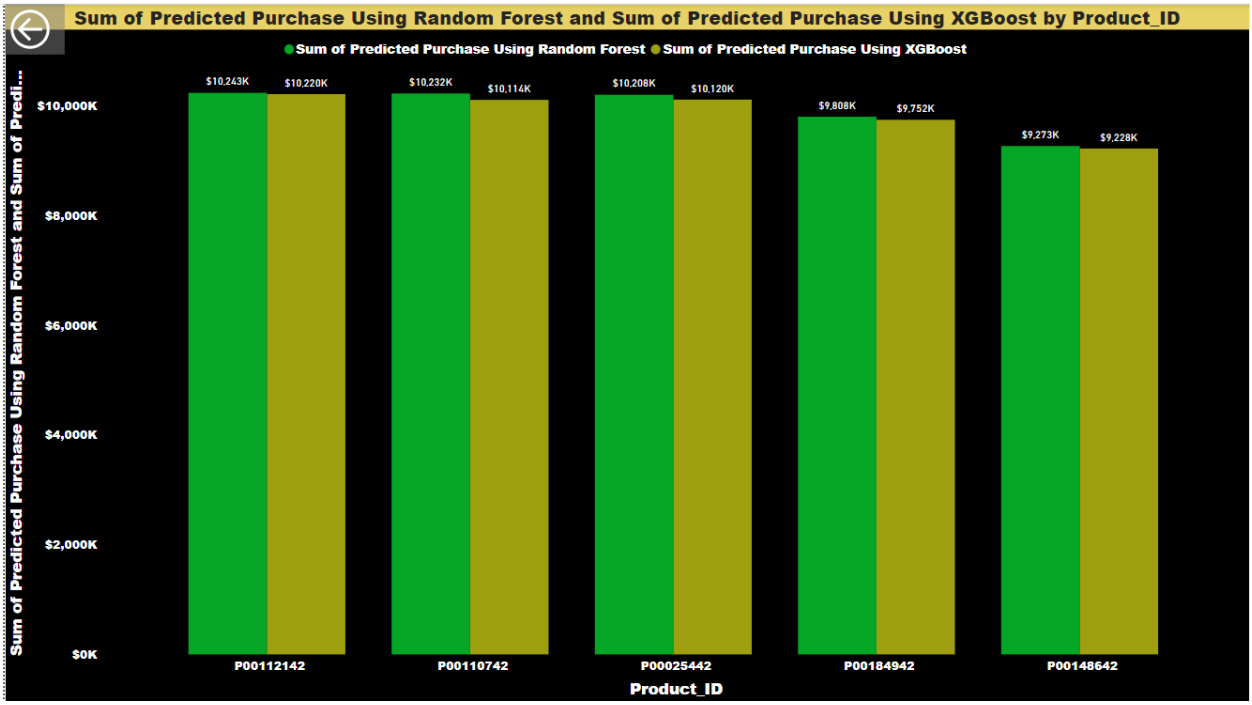
(9) <u>Sum of Predicted Purchase by Product Category 1,2 and 3 in Test Dataset for XGBoost</u>



From the Above Insight, following data can be seen:

(a) Sum of Predicted Purchase is highest in Product category 1 at $6,71,825k for Male in 1 and $1,44,444k for female in 1.
(b) Sum of Predicted Purchase is highest in Product category 2 at $6,24,094k for Male in 8 and $2,07,438k for female in 8.
(c) Sum of Predicted Purchase is highest in Product category 3 at $11,48,937k for Male in 16 and $3,61,033k for female in 16.

(10) **Sum of Predicted Purchase Using Random Forest and XGBoost for Product id in Test Dataset**



From the Above Insight, following data can be seen:

(a) Sum of Predicted purchase using Random forest regressor in Test Data is highest with $10,243k in Product id P00112142.

(b) Sum of Predicted purchase using XGBoost regressor in Test Data is highest with $10,220k in Product id P00112142.