Exploratory Data Analysis

# **UNIVARIATE ANALYSIS**

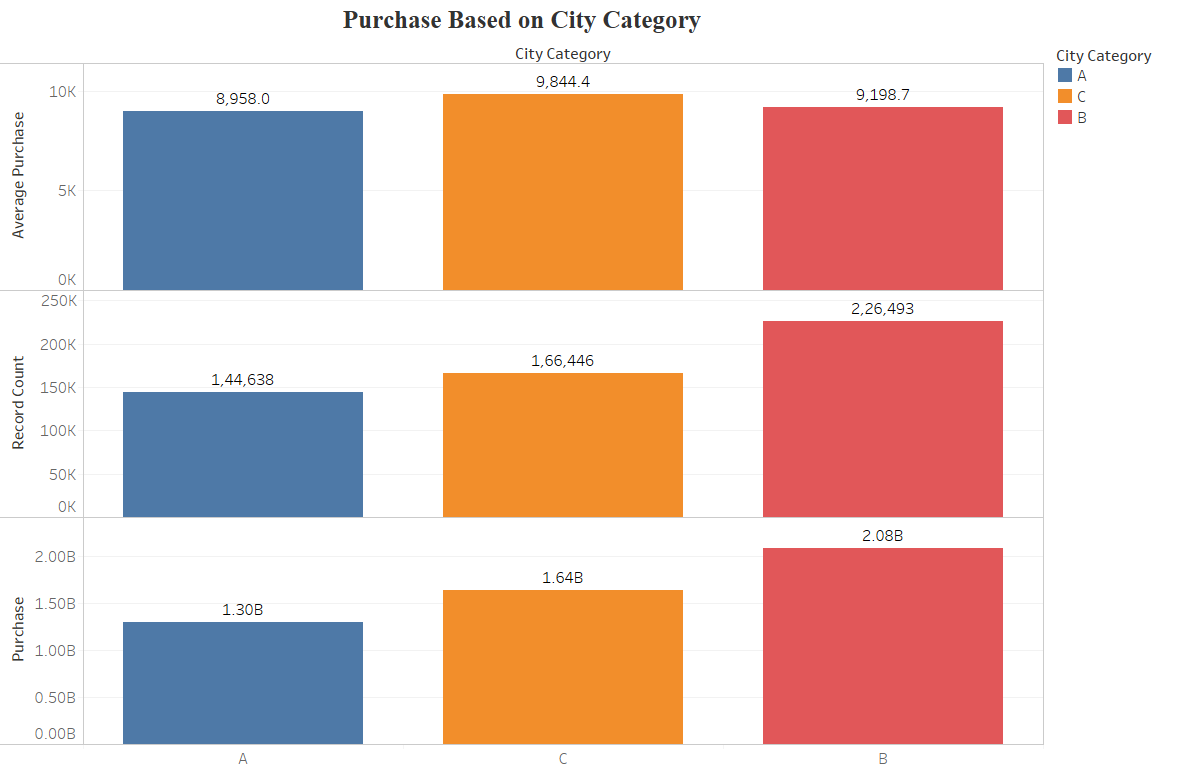
## **AGE**



**Observations:**

* 40% of the shopping i.e., almost 2 Billion $ is spent by the age group **’26-35’**.
* **‘0-17’** age group account for the least Sales i.e., 132 Million $ only.
* Interestingly, all the age groups spend on an average approx. 9000 $

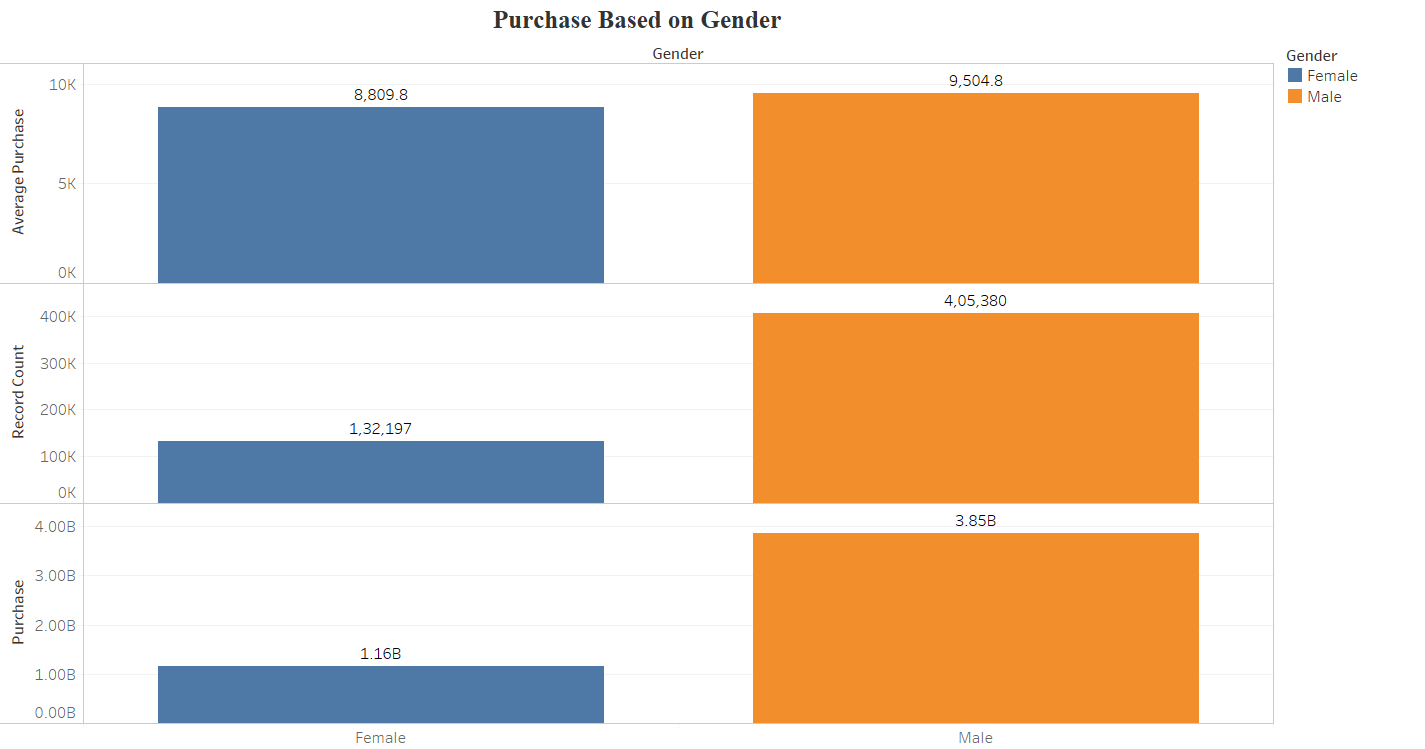
## **CITY – CATEGORY**



**Observations:**

* City-C accounts for highest Purchase and City-A for the lowest amongst all the cities.
* City-B on an average spend 9844

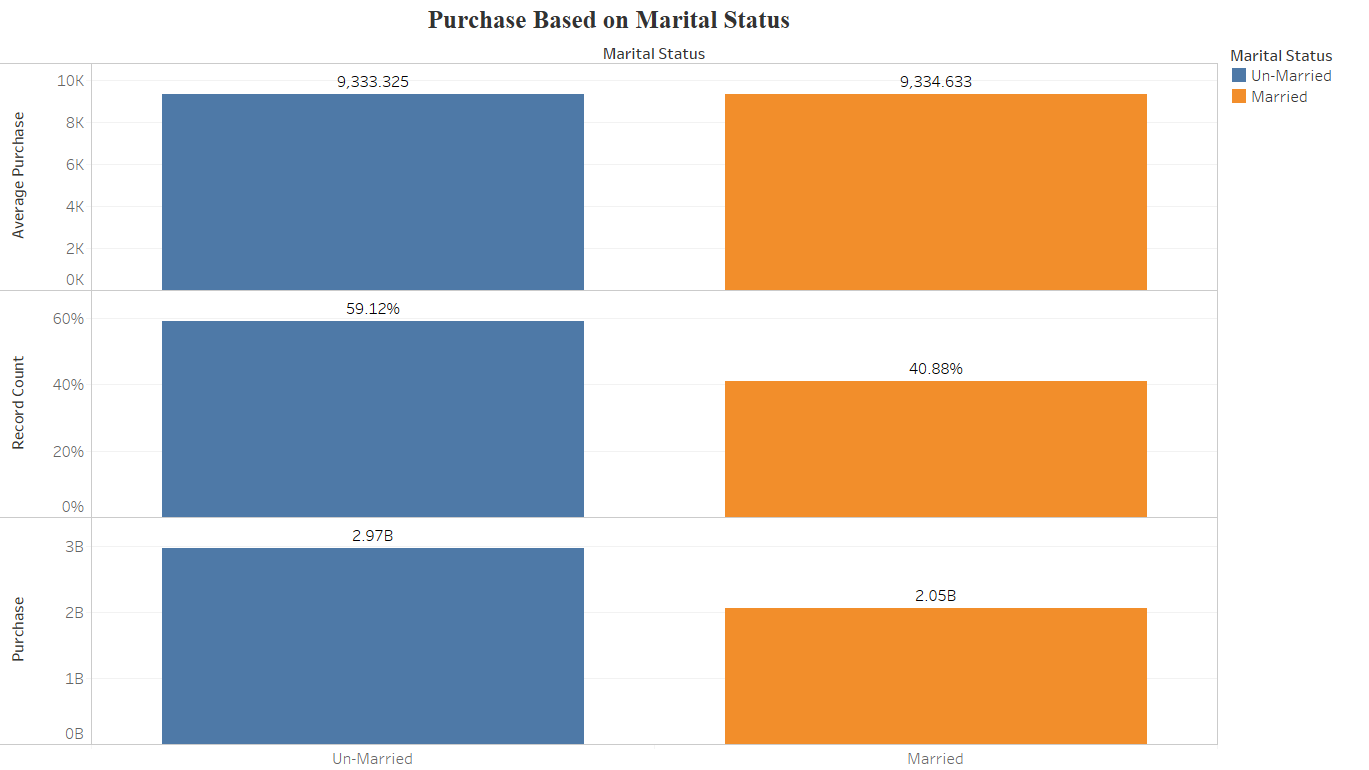
## **GENDER**



**Observations:**

* Male Shopper account for almost 75% of the Purchases both in terms of Frequency and Spending.
* Male Shoppers on an average spend almost 700$ more than Female Shoppers.

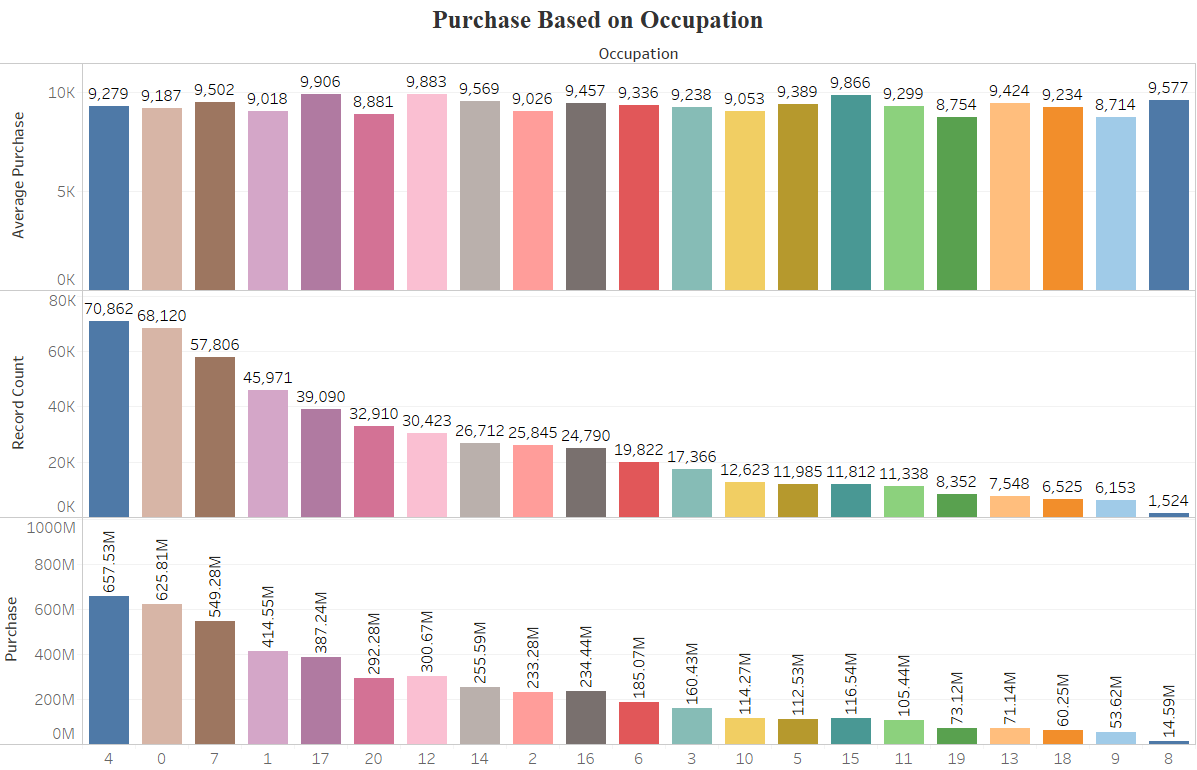
## **MARITAL STATUS**



**Observations:**

* Unmarried people account for an excess of 1 B$ Purchase than Married people.
* Interestingly, Average Purchase committed by Married and Unmarried individuals is the same i.e., approx. 9333$

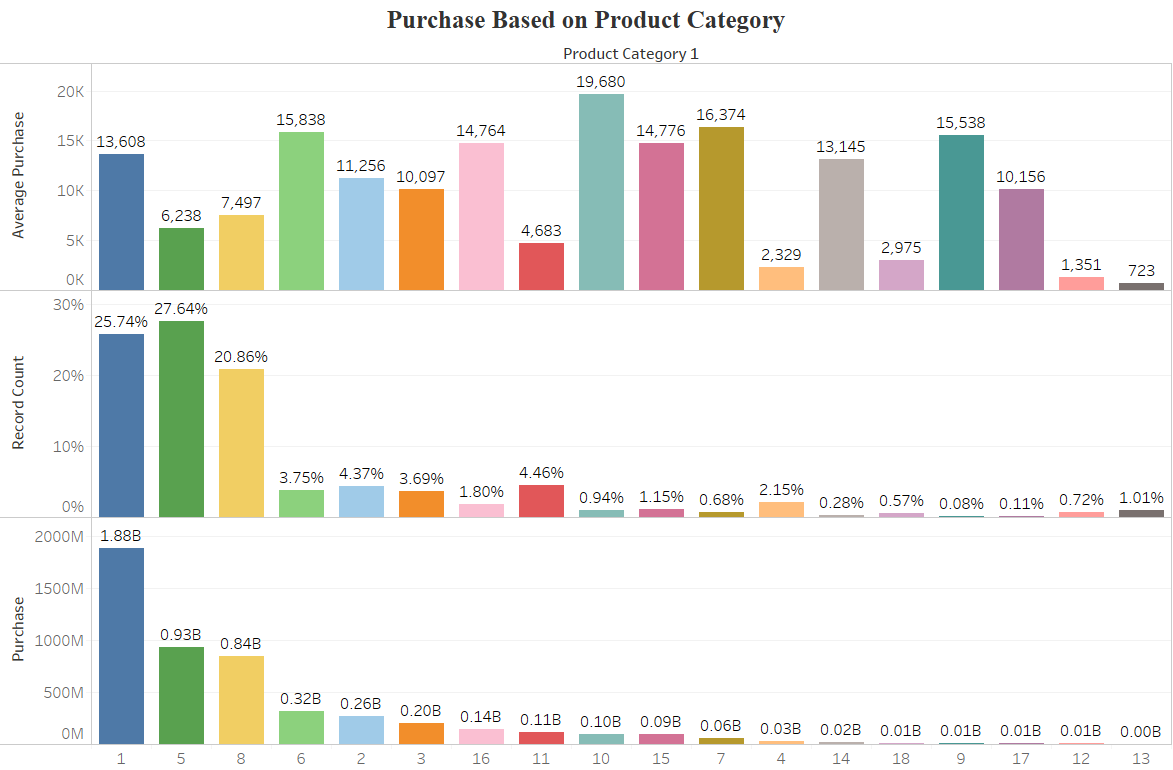
## **OCCUPATION**



**Observations:**

* Occupation – **4,0 and 7** are responsible for purchase above 500 M$ each.
* There are 13 Occupation Categories within which Customers spend between 100-500 M$.
* Occupation – **19,13,18,9 and 8** in total account for less than 100 M$ Purchase.

## **PRODUCT-CATEGORY**



**Observations:**

* Product Category -1 accounts for highest Revenue and is the 2nd most popular product in terms of Sales. Also, Customers spend 4275$ more on it than Average Purchase.
* Product Category-5 is the most popular Product in terms of Sales.
* On Average, Customers are ready to spend the most on Product Category-10.
* Product Category – 12 and 13 have Low Sales and Lowest Average Spending.

**Business Application:**

* Product Category – **5,1,and 8** account for 75% of Purchase by the Customers:
  + If the Price of **‘Product-5’** is increased by **5%**, **‘Product-1’** by **3%** and **‘Product-8’** by **4%** then the Revenue will increase by 150 M$ which is higher than the combined Revenue of eight lowest selling Products.

# **BIVARIATE ANALYSIS**

## **Age – City Category**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **City Category** | | | | | |
|  | **A** | | **B** | | **C** | |
| **Age** | **% City Population** | **Avg. Purchase** | **% City Population** | **Avg. Purchase** | **% City Population** | **Avg. Purchase** |
| 0-17 | 1.73% | 8,673 | 2.33% | 8,985 | 4.16% | 9,172 |
| 18-25 | 18.68% | 8,887 | 18.75% | 9,070 | 16.91% | 9,820 |
| 26-35 | 49.81% | 8,990 | 39.63% | 9,199 | 31.77% | 9,953 |
| 36-45 | 18.07% | 9,042 | 20.58% | 9,150 | 20.88% | 10,009 |
| 46-50 | 5.16% | 8,386 | 8.79% | 9,297 | 10.31% | 9,662 |
| 51-55 | 4.13% | 9,575 | 7.70% | 9,394 | 8.54% | 9,918 |
| 55+ | 2.41% | 8,587 | 2.22% | 9,886 | 7.44% | 9,523 |

**Observations:**

* Age group **26-35**  is the most frequent shopper group.
* In City-C, except Teens all the Age Group spend above the Average.
* The Age Groups **’18-45’** account for 70% and above Product shopping in all the cities.

## **Age – Gender**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Gender** | | | |
|  | **Female** | | **Male** | |
| **Age** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** |
| 0-17 | 3.75% | 8,445 | 2.41% | 9,312 |
| 18-25 | 18.20% | 8,405 | 18.15% | 9,507 |
| 26-35 | 37.33% | 8,792 | 40.79% | 9,471 |
| 36-45 | 19.99% | 9,047 | 20.00% | 9,517 |
| 46-50 | 9.72% | 8,929 | 7.81% | 9,429 |
| 51-55 | 7.29% | 9,131 | 6.90% | 9,789 |
| 55+ | 3.73% | 9,120 | 3.94% | 9,557 |

**Observations:**

* Male Customers on an average spend more than Female Customers.
* Teenage Girls spend less than older Women.

## **Age – Marital Status**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Marital Status** | | | |
|  | **Un-Married** | | **Married** | |
| **Age** | **% Shopper** | **Avg. Purchase** | **% Shopper** | **Avg. Purchase** |
| 0-17 | 4.63% | 9,020 |  |  |
| 18-25 | 24.23% | 9,281 | 9.39% | 9,065 |
| 26-35 | 41.07% | 9,313 | 38.30% | 9,316 |
| 36-45 | 20.45% | 9,470 | 19.34% | 9,296 |
| 46-50 | 3.88% | 9,035 | 14.65% | 9,381 |
| 51-55 | 3.35% | 9,664 | 12.28% | 9,604 |
| 55+ | 2.40% | 9,660 | 6.04% | 9,335 |

**Observations:**

* Older people are likely to spend more irrespective of their marital status.

## **City – Gender**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Gender** | | | |
|  | **Female** | | **Male** | |
| **City Category** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** |
| A | 26.33% | 8,631 | 27.09% | 9,062 |
| B | 42.73% | 8,591 | 41.94% | 9,401 |
| C | 30.94% | 9,265 | 30.97% | 10,033 |

**Observations:**

* Male Customers : *City C* customers are likely to spend the *most*, *City A* customers are likely to spend the *least*.
* Female Customers: *City B* are likely to spend the *least* whereas *City C* the *most*.

## **City – Marital Status**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Marital Status** | | | |
|  | **Un-Married** | | **Married** | |
| **City Category** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** |
| A | 28.13% | 9,021 | 25.14% | 8,856 |
| B | 42.23% | 9,192 | 41.99% | 9,208 |
| C | 29.64% | 9,830 | 32.87% | 9,863 |

**Observations:**

* Irrespective of Marital Status, People in City C spend the most and City A the least.

## **CITY – OCCUPATION**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **City Category** | | | | | |
|  | **A** | | **B** | | **C** | |
| **Occupation** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** |
| 0 | 12.72% | 8,919 | 12.96% | 9,078 | 12.24% | 9,586 |
| 1 | 8.57% | 8,778 | 8.52% | 8,802 | 8.57% | 9,517 |
| 2 | 6.15% | 8,588 | 4.90% | 8,847 | 3.51% | 10,032 |
| 3 | 3.84% | 8,821 | 3.02% | 9,281 | 2.98% | 9,645 |
| 4 | 16.50% | 8,989 | 12.94% | 9,107 | 10.63% | 9,954 |
| 5 | 1.62% | 8,937 | 2.88% | 9,132 | 1.87% | 10,266 |
| 6 | 2.54% | 9,302 | 4.72% | 9,152 | 3.29% | 9,720 |
| 7 | 10.93% | 8,884 | 9.98% | 9,442 | 11.65% | 10,076 |
| 8 | 0.07% | 11,443 | 0.37% | 9,480 | 0.36% | 9,397 |
| 9 | 0.48% | 8,985 | 1.40% | 8,581 | 1.37% | 8,818 |
| 10 | 1.51% | 8,662 | 1.82% | 8,953 | 3.79% | 9,254 |
| 11 | 1.66% | 9,382 | 2.42% | 8,704 | 2.07% | 10,190 |
| 12 | 4.79% | 9,496 | 5.68% | 9,866 | 6.39% | 10,155 |
| 13 | 0.27% | 8,684 | 1.03% | 8,844 | 2.90% | 9,763 |
| 14 | 5.27% | 9,483 | 4.88% | 9,323 | 4.83% | 9,987 |
| 15 | 2.07% | 9,807 | 2.24% | 9,788 | 2.25% | 10,019 |
| 16 | 4.09% | 9,450 | 4.71% | 9,201 | 4.93% | 9,794 |
| 17 | 5.57% | 9,140 | 6.93% | 9,814 | 9.22% | 10,402 |
| 18 | 1.10% | 8,871 | 0.92% | 8,817 | 1.70% | 9,745 |
| 19 | 1.53% | 8,362 | 1.49% | 8,900 | 1.67% | 8,890 |
| 20 | 8.72% | 8,361 | 6.19% | 9,064 | 3.77% | 9,518 |

**Observations:**

* In City-C, 18 different Occupation type People spend above average i.e., more than 93%
* In City-A and City-B, not more than 6 types of Occupation Type are observed to spend above Average.

## **CITY – STAY**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **City Category** | | | | | |
| **A** | | **B** | | **C** | |
| **Stay In Current City Years** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** | **% Shoppers** | **Avg. Purchase** |
| 0 | 16.39% | 9,029 | 12.43% | 8,947 | 12.55% | 9,900 |
| 1 | 33.30% | 8,928 | 36.04% | 9,225 | 35.69% | 9,768 |
| 2 | 18.35% | 9,015 | 18.01% | 9,251 | 19.29% | 9,900 |
| 3 | 16.86% | 8,957 | 18.53% | 9,231 | 16.20% | 9,893 |
| 4+ | 15.10% | 8,880 | 15.00% | 9,240 | 16.27% | 9,856 |

**Observations:**

* Customers who arrived recently in City-B and City-C shop less frequently than those who stayed longer (Acclimatization can be an issue).

## **GENDER-OCCUPATION**

**Observations:**

* Female Customers of Occupation category – **0,1 and 4** are likely to shop the most.
* Female Customers of Occupation category – **8,15,17,18** are likely to spend above average.
* Male Customers of Occupation category – **0,4 and 7** are likely to shop the most.
* Male Customers of Occupation category – **4,5,6,7,8,10,12,13,14,15,16 and 17** are likely to spend above average.

## **MARITAL STATUS – OCCUPATION**

**Observations:**

* On an Average, Unmarried people of Occupation – **9** spend the least and of Occupation – 15 spend the most.
* On an Average, Married people of Occupation – **2** spend the least and of Occupation – 17 spend the most.

## **OCCUPATION – STAY**

**Observations:**

* People who came recently to a City are likely to spend more on Purchase than People who have been residing in the City for more than a Year.
* Occupation – **0,1,4 and 7** are the most frequent Shoppers, irrespective of duration of Stay in a City.

## **PRODUCT**

**Observations:**

* Irrespective of any other variable:
  + Product Category – **1,2,3,6,7,9,10,14,15** and **16** record an above Average spending by the Customers.
  + Product Category –**4,5,8,11,12,13** and **18** record a below Average spending by the Customers.
  + Product Category – **1,5,** and **8** are the most popular products i.e., these Products are bought more frequently by the Customers.
  + Product Category – **9,14, 17**and **18** are the least popular products i.e., these Products are bought less frequently by the Customers.

# **STATISTICAL ANALYSIS**

## **CHI-SQUARE TEST OF INDEPENDENCE**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AGE** | **CITY CATEGORY** | **GENDER** | **MARITAL STATUS** | **OCCUPATION** | **PRODUCT CATEGORY-1** | **STAY** |
| **AGE** | YES | YES | YES | YES | YES | YES | YES |
| **CITY CATEGORY** | YES | YES | YES | YES | YES | YES | YES |
| **GENDER** | YES | YES | YES | YES | YES | YES | YES |
| **MARITAL STATUS** | YES | YES | YES | YES | YES | YES | YES |
| **OCCUPATION** | YES | YES | YES | YES | YES | YES | YES |
| **PRODUCT CATEGORY-1** | YES | YES | YES | YES | YES | YES | YES |
| **STAY** | YES | YES | YES | YES | YES | YES | YES |