Exploratory Data Analysis Financial Institute

EDA

- Data Scientists are focused to normalize and analyze data statistically, categorically using plots to find potential benefits for an organization.
- Exploratory Data Analysis popularly known as EDA is a process of performing some initial investigations on the dataset to discover the structure and the content of the given dataset.
- It is often known as Data Profiling. It is an unavoidable step in the entire journey of data analysis right from the business understanding part to the deployment of the models created.
- EDA is where we get the basic understanding of the data in hand which then helps us in the further process of Data Cleaning & Data Preparation.
- Analyze Database in visualize way that it shows the correlation and impact on each other.
- Understand Database/variables. Find missing values.
- Find outliers to get better analytics results.
- Find what are the different variables in database that influence the ask.
- Find what are the factors affect the ask.
- Fine what could be plausible reason for that.
- Find how the data is distributed.

Financial Institute Background & Goal

- A Financial institute would like to generate more profit; Hence producing new service of Trading Facilitation Service. However, this service needs huge initial investment in Infrastructure set-up, Licensing and additional Employee cost.
- To cover-up this cost and gain profit, they are planning to put percentage-based commission on every trade. However, considering its commonly available feature in competitive institutions, they want to implement an offer of some Discounts to whole customer base.
- When you offer some discount to whole customer base, real problem is that it hampers the profit heavily due to many customers do not trade in large quantities. So they want to offer Discount selectively.
- The criteria for offering discount should be those customers who are Trading Heavily. This analysis will help them to select customers who get discounts.

FINANCIAL INSTITUTION'S GOAL

- Make more money while customer do more trading.
- Retain & Attract more customers.

Data Scientist Approach

- We are looking into and analyzing customer's who are trading heavily. That means, our focus will be on in comparison of
  - No of Children vs Derivative Value
  - Family's financial income vs Derivative Value
  - Customer Occupation and Customer's Partner's Occupation vs Derivative Value
  - Total Investment (like in Mutual Fund, Tax Saving Bonds, Equity, Commodity) vs Derivative Value
  - Total Expense (like Life-Medical-Term Insurance, Home-Personal Loan, Card-Online Purchase) vs Derivative Value etc

Data Scientist ASSUMPTIONs

- Considering we don't have direct interaction with Business team or Product Managers, so based on what we understand, we have built this analysis.
    - We are looking for those who are doing TRADING i.e. Investment in Derivative variable to be focused.
    - REF_NO is just a serial number, so ignoring it right now. In further analysis, if business wants list of those customers, we can provide based on REF_NO.
    - YEAR_LAST_MOVED is not understood correctly and finding no relation with what data we have. If this field purpose is to show since how many year customer is with Financial Institute, then we can analyze it for retention purpose. So ignored as of now.
    - POST_CODE / POST_AREA are ignored considering we do have REGION variable already.
    - REVENUE_GRID is field will be updated post analysis by Financial Institute as per the understanding. So ignored as of now.
    - INVESTMETNS like Investment in Mutual Fund, Investment in Tax Saving Bonds, Investment in Equity, Investment in Commodity are signs of Wealth of Customer, but they are focused on Investments; not Trading.
    - EXPENSE like all kind of Insurance, Loans, Online Purchase or Credit Card Bills are necessity in today's time but that is negative signs of customer as if it's high.
    - There is no Exact Understanding of variables in relevance with Derivative/Trading capability/potential. So, every observation, conclusion and recommendation are based on what we can think/interpret.

Data Information

Variable	Description
REF_NO	Serial Number of Record
children	Number of Children at home
age_band	Age band/group of the customer
status	Marital Status of customer
occupation	Occupation of Customer
occupation_partner	Occupation of Customer's Partner
home_status	Status of home living in
family_income	Family Income Range
self_employed	Whether self employed or not (Yes/No)
self_employed_partner	Whether partner is self employed or not (Yes/No)
year_last_moved	Year when last moved to current location
TVarea	TV News Followed
post_code	Postal Code of the customer
post_area	Postal Area of the customer
Average Credit Card Transaction	Average Credit Card Transaction per month
Balance Transfer	Balance Transfer per month
Term Deposit	Term Deposit ?
Life Insurance	Life Insurance Premium
Medical Insurance	Medical Insurance ?
Average A/C Balance	Average Account Balance of customer
Personal Loan	Personal Loan Remaining Value
Investment in Mutual Fund	Customer's investment in Mutual Funds
Investment Tax Saving Bond	Customer's Investment in Tax Saving Bonds
Home Loan	Home Loan EMI
Online Purchase Amount	Customer's Online Purchase Capability
Revenue Grid	Revenue Grid - 1 is Revenue Generator and 2 is not
gender	Customer is Male or Female
region	Which region of states customer is living
Investment in Commodity	Investment in Commodity
Investment in Equity	Investment in Equity
Investment in Derivative	Investment in Derivative
Portfolio Balance	Customer's Portfolio Balance Today

Data Information

10155 Observations with 32 Variables (columns).

18 Integers and 14 Object Variables with storage of 2.5+mb.

Converting 14 Object to Categorical variable, storage reduced to 2mb.

Children, Age_Band, Status, Occupation, Occupation_Partner, Home Status, Family_Income, Self_employed, Self_employed_partner, Tvarea, Post_Code, Region are Categorical variables.

Average Credit Card Transaction, Term Deposit, Life Insurance, Medical Insurance, Average A/C Balance, Personal Loan, Investment in Mutual Fund, Investment Tax Saving Bond, Home Loan, Online Purchase Amount, Investment in Commodity, Investment in Equity, Investment in Derivative, Portfolio Balance are integer variables.

Integer variables are assumed as Expense (routine outflow of money) and Investment (for future).

Expenses: Average Credit Card Transaction, Term Deposit, Life Insurannce, Medical Insurannce Average A/C Balance, Personal Loan, Home Loan, Online Purchase Amount

Investments: Investment in Mutual Fund, Investment Tax Saving Bond, Investment in Commudity, Investment in Equity, Portfolio Balance

Investment in Derivative is TRADING.

Data Information

Observations On Data

- Children
  - Customer database shows 5 unique values in children variable (0,1,2,3,4+).
  - 6208 Customer's having 0 children, maximum occurrence

- Age_Band
  - There are 13 different values in Age_Band variable
  - 1359 times Range 45-50, maximum occurrence

- Status
  - There are 5 different values in Status variable
  - 7709 times Partner, maximum occurrence

- Occupation / Occupation_Partner
  - There are 9 different values in Occupation/Occupation Partner
  - 2449 times Professional in Occupation and 2394 times Unknown in Occupation Partner

- Home Status
  - There are 5 unique values in Home Status of customer database
  - 9413 customers have their own home

- Family Income
  - There are 13 unique values/ranges of Family Income of customer
  - 2517 customers have >=35000 family income as per customer database

Data Information

- Self Employed/ Self Employed Partner
  - It's only Yes or No value column
  - 9436 customer are not Self employed whereas 9026 Partner's are not Self Employed
- TVArea
  - Customer watches / following 14 unique TV Channels
  - Central is followed by 1618 Customers, maximum for any channel
- Gender
  - Customer classified in three categories like Female, Male and Unknown
  - 7634 customers are female out of 10155 which is more than 70% of total
- Region
  - Customer classified 13 different/unique regions
  - 2100 customers are from Southeast region

Python Libraries Used

- numpy
  - Numerical Python (numpy) is used for working with multi-dimensional arrays.
- Pandas
  - Pandas is used for data manipulation and analysis.It offers structures for data manipulations.
- matplotlib.pyplot
  - Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
  - matplotlib.pyplot is a state-based interface to matplotlib. Pyplot is mainly intended for interactive plots.
- %matplotlib inline
  - Displays output inline. IPython kernel has the ability to display plots by executing code.
- Seaborn –
  - Seaborn is a Python data visualization library built on top of Matplotlib.
  - Seaborn contains a number of patterns and plots for data visualization.

Univariate Analysis
Financial Institute

EDA – Investment in Derivative

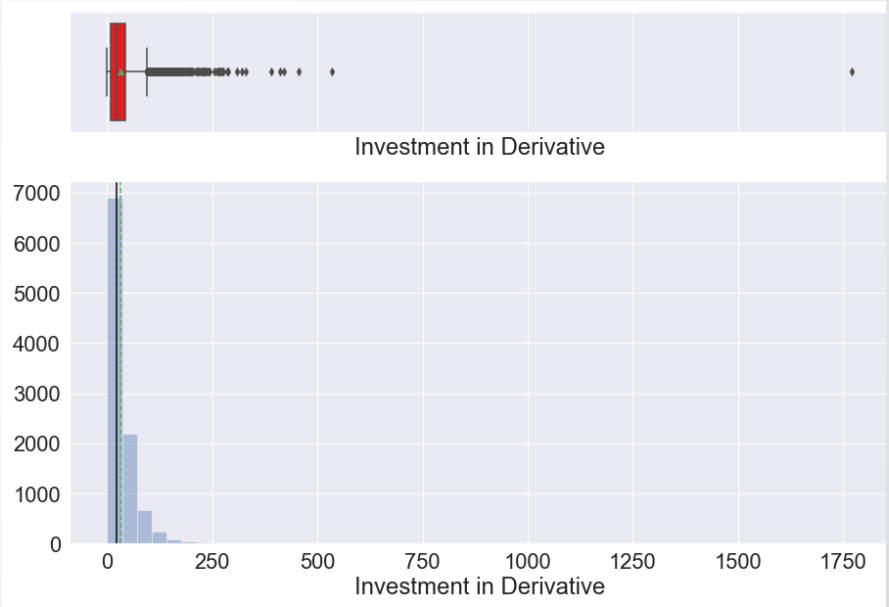
Investment In Derivative/Trading:

- Commission gets generated from Derivatives/Trading heavily. And Derivative can be done by those who are Risk Takers or having high Risk Appetite.

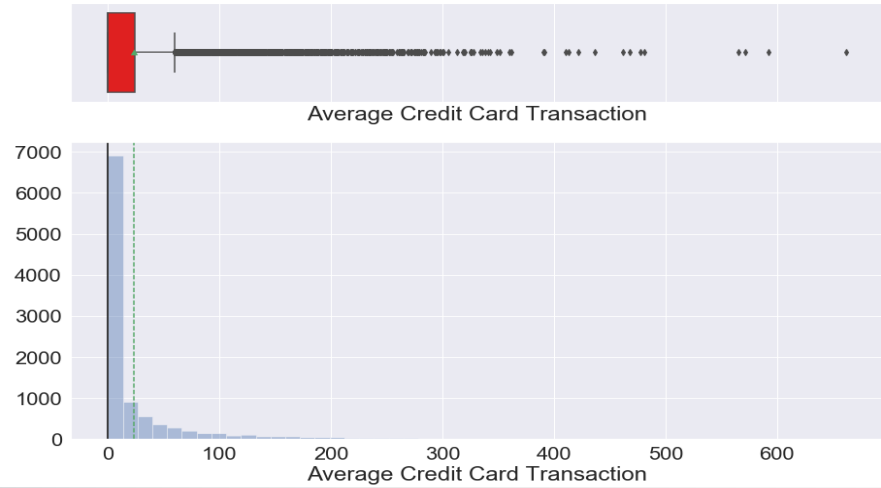
- Let's explore.

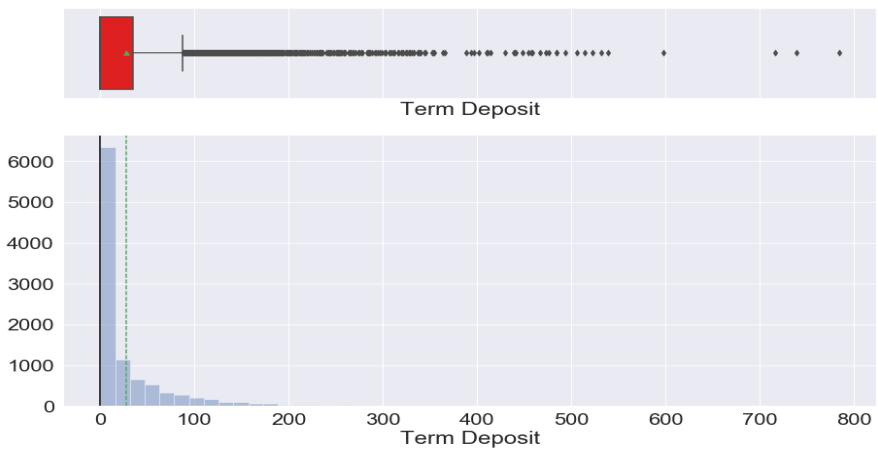
- Derivative data is skewed right side based on Box Plot.
- Majority (more than 7000) TRADING values are almost less than MEAN (close to 32)
- Lots of outliers
- While mostly TRADING are less than or close to 105; we have an observation to have very high as well as like between 100 to 250 and 250 to 1771 as seen in Box plot here

- Overall, it shows there is big scope and potential for Derivatives



EDA – ACCT and Term Deposit





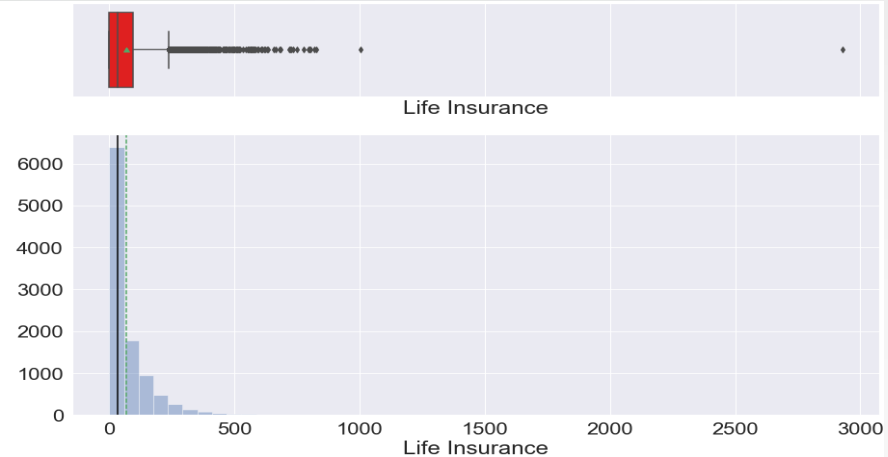
Average Credit Card Transaction:

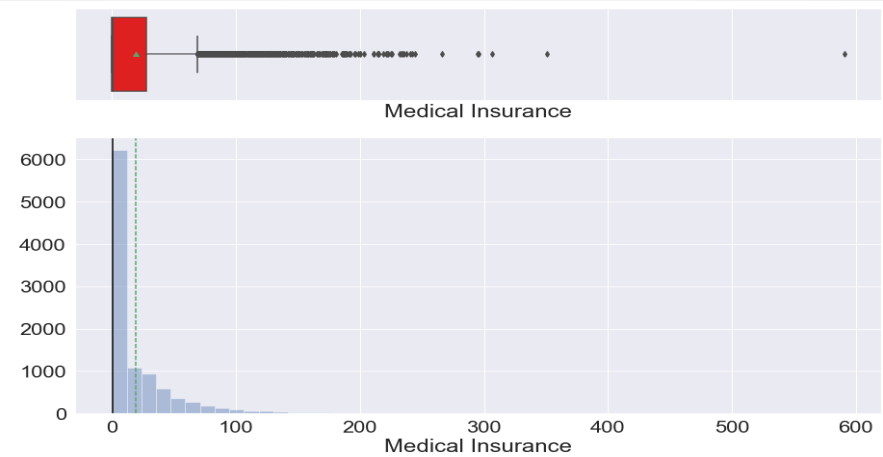
- ACCT data is Highly Skewed Data (3+ value)
- ACCT is skewed right side based on Box Plot
- Majority ACCT are close to 0 or within MEAN point
- STD is just more than double of MEAN
- ACCT is normal showing mature customers who knows impact of high credit card transaction impacts to their financial specimens.

Term Deposit:

- Term Deposit is skewed right side based on Box Plot
- Majority (more than 7500) Term Deposits are close to or less than MEAN
- 75% Term Deposits are less than STD (~35)
- Term Deposit is expense for must do kind of. Based on data it looks like not having any impact on customer potential with respect to Derivative.

EDA – Life & Medical Insurance





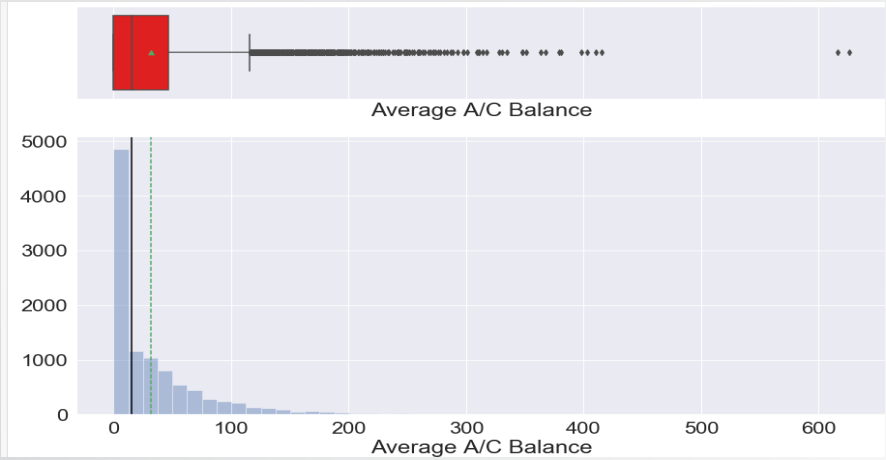
Life Insurance:

- Life Insurance data is Highly Skewed Data (~5 value) & is skewed right side based on Box Plot
- Majority (more than 6000) Life Insurance values are less than MEAN and nearly 2000 Life Insurance values are in range of 60 to 120
- Overall, 75% Life Insurance values are within STD value
- Life Insurance is normal showing mature customers who knows impact value of having Life Insurance. Based on data it looks like majorly it's normal premium they are paying which should not have any impact on their Derivatives.

Medical Insurance:

- Medical Insurance data is Highly Skewed Data (3+ value) & is skewed right side based on Box Plot
- Majority (more than 7000) Medical Insurance values are less than MEAN and nearly 2000 Medical Insurance values are in range of 12+ to 36
- Overall, 75% Medical Insurance values are way less than or close to STD value
- Medical Insurance is normal showing mature customers who knows it's must have it in today's time. Based on data it looks like majorly it's normal premium they are paying which should not have any impact on their Derivatives.

EDA – AACB and Personal Loan



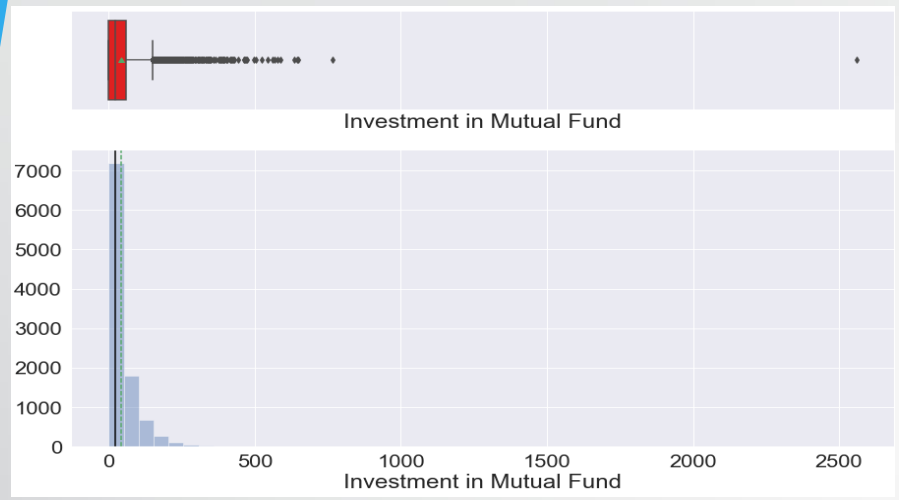
Average A/c Balance:

- Highly Skewed Data (2+ value)
- AACB is skewed right side based on Box Plot
- Majority (more than 7000) AACB values are less than MEAN
- Nearly 2000 AACB values are in range of 24+ to 48
- Overall, 85% AACB values are way less than or close to STD value
- AACB data shows Average A/C Balance is not healthy post all expense and investments. That may be a Negative for attracting such customer for Trading/Derivative.

Personal Loan:

- Personal Loan shows Highly Skewed Data (23+ value)
- Personal Loan data is skewed right side based on Box Plot
- Majority (more than 9000) Personal Loan values are almost ZERO or less than MEAN
- Based on data it looks like most of the customers do not carry any Personal Loan which is positive as they don't carry any financial burden

EDA – Investment In MF & Tax Saving Bond



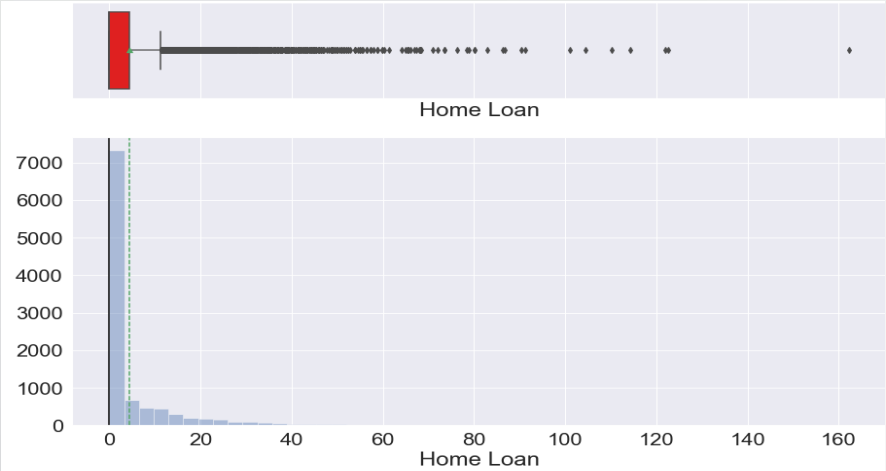
Investment In Mutual Fund:

- IMF shows Highly Skewed Data (8++ value)
- IMF data is skewed right side based on Box Plot
- Majority (more than 7000) IMF values are close to ZERO or less than MEAN
- Nearly 2000 IMF values are in range of 50 to 100
- IMF data shows customer is very disciplined and understands value of long-term investment for good returns with small value

Investment In Tax Saving Bonds:

- ITSB showing Highly Skewed Data (3+ value)
- ITSB data is skewed right side based on Box Plot
- Majority (more than 7000) IMF values are close to ZERO or less than 3.5
- Mostly ITSB are less than or close to 20
- ITSB Data shows customer not much interested in Tax Saving Bonds rather preferring direct investments

EDA – Home Loan and Online Purchase



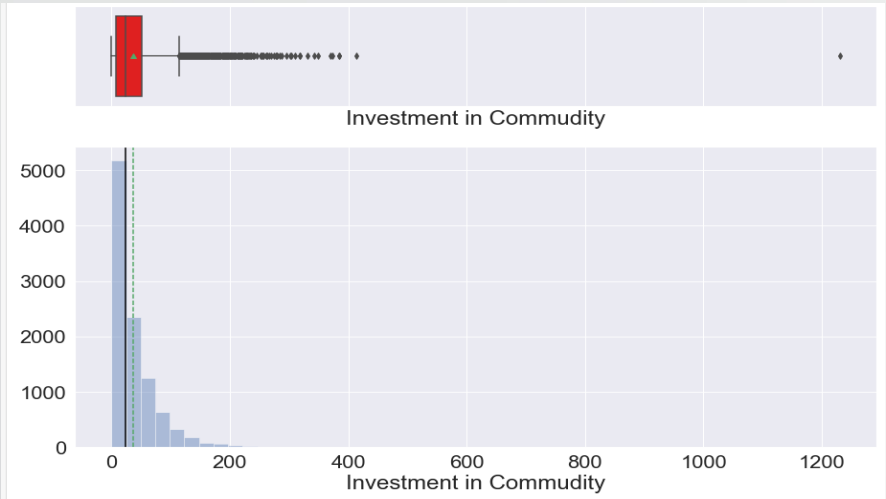
Home Loan:

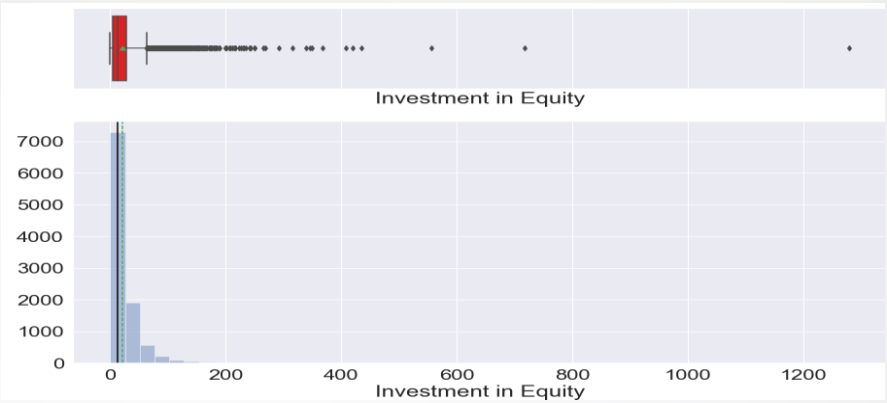
- Home Loan data shows Highly Skewed (3+ value) and data is skewed right side based on Box Plot
- Majority (more than 7000) Home Loan values are close to ZERO and while mostly Home Loan are less than or close to or less than 5 (MEAN); we have an observation to have very high Home Loan value (above 20) as well as like max is 162+
- Home Loan data shows one positive observation that 7000+ do not have any Home Loan Burden and may be they are within owning homes and rest have lessor EMI on the monthly expenses

Online Purchase Amount:

- Online Purchase Data shows Highly Skewed (20+ value)
- OPA data is skewed right side based on Box Plot
- Majority (more than 9500) OPA values are less than 100
- OPA Data shows most of the customers have habit but disciplined in shopping

EDA – Investment in Commodity & Equity





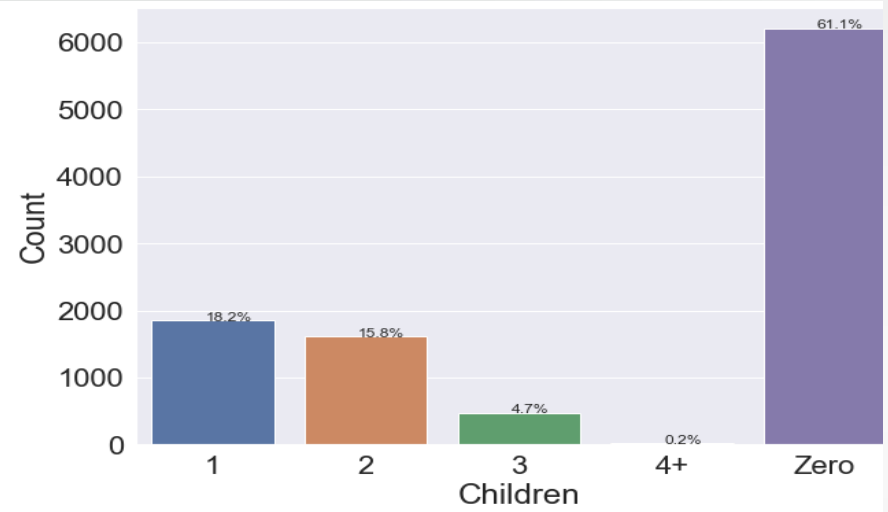
Investment in Commodity:

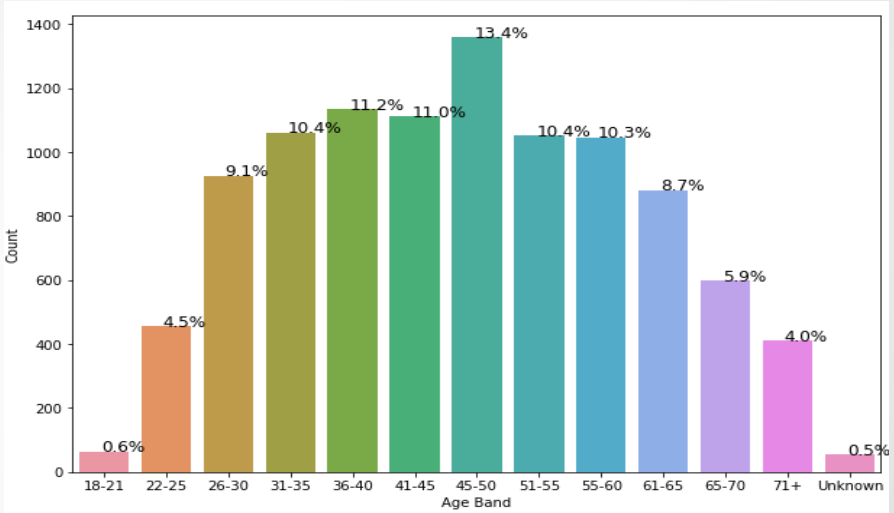
- ICOM data shows Highly Skewed (4+ value)
- ICOM data is skewed right side based on Box Plot
- Majority (more than 5000) ICOM values are less than MEDIAN
- More than 2000 ICOM values are in range of 25 to 51 (75%tile)
- ICOM data shows customers are disciplined and does invest smartly with diversification of Equity, MF, Commodity etc

Investment in Equity:

- IEQ data shows Highly Skewed Data (10+ value)
- IEQ data is skewed right side based on Box Plot
- Majority (more than 7000) IEQ values are close to or less than MEAN (25)
- Nearly 2000 IEQ values are in range of 25 to 50
- IEQ data shows customers are disciplined and does invest smartly with diversification of Equity, MF, Commodity etc

EDA – Children & Age Band





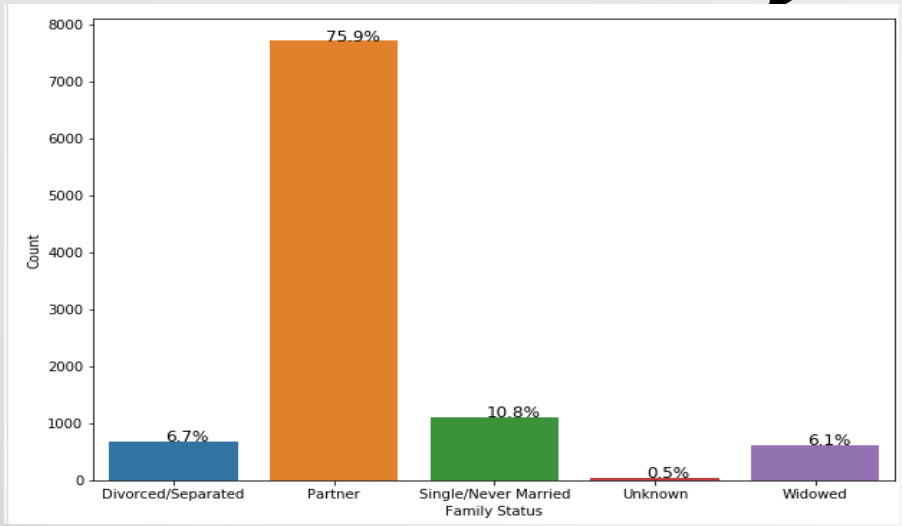
Children:

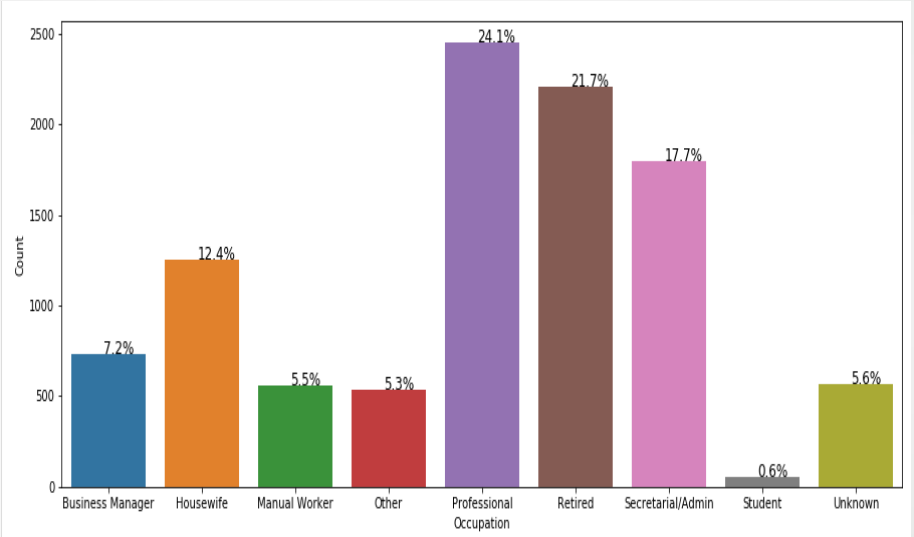
- Nearly 61% ; 6200+ customers do not have kids/children.
- Nearly 18% have only 1 child
- Based on data, we can assume/predict that almost 8000 Customers having lessor expense compared to others. So, they are potential customers to get attracted towards Derivative

Age_Band:

- 13+% ; 1350+ customers are in 45-50 years age group/band, leading; should be at peak of their career.
- All other customers are distributed nicely in other Age Groups.
- Mostly Derivative/Trading is attracted by those who are settled in their career.

EDA – Family Status and Occupation





Family Status:

- Almost 76% ; 7700+ customers are living as Partners, leading population.
- All other customers are distributed as
  - Divorced/Separated (6.7%)
  - Single/Never Married (10.8%)
  - Widowed (6.1%) and 0.5% are unknown.
- We can't assume that individual living with Partner can do more Trading. But we can predict that individual living with Partner/Spouse are matured Customers who always look for opportunities to earn something Extra.

Occupation:

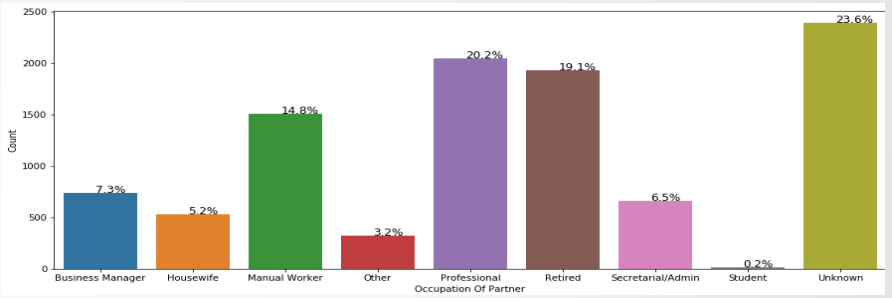
- Including Housewives, there are 7 different Occupations; and Others along with Unknowns.
- Nearly 24+% ; 2450 customers are Professionals, leading population.
- Next to Professionals, nearly 22% customers are retired and followed by nearly 18% Secretarial/Admin (gov) jobs.
- There are nearly 10+% customer's occupation is unknown/.others.
- Professionals, Retired and Gov Job having good worth and can be attracted by Derivatives.

EDA – Multiple Variables

Partner's Occupation:

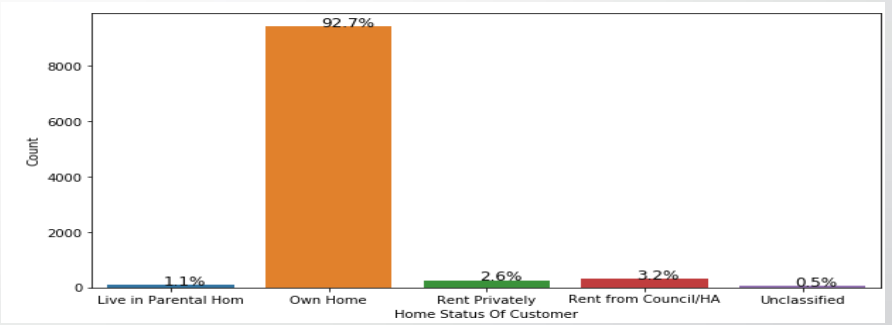
- Key observation to make a note is there are 23.6% i.e 2394 customers partner's occupation is unknown; followed by 20.2% customers' partners are Professionals and 19.1% are Retired.

- Where both, Customer and Partner, are Professional or Retired or Gov job, it's high probability they get into Trading easily. It's big Opportunity.



Home Status Of Customer:

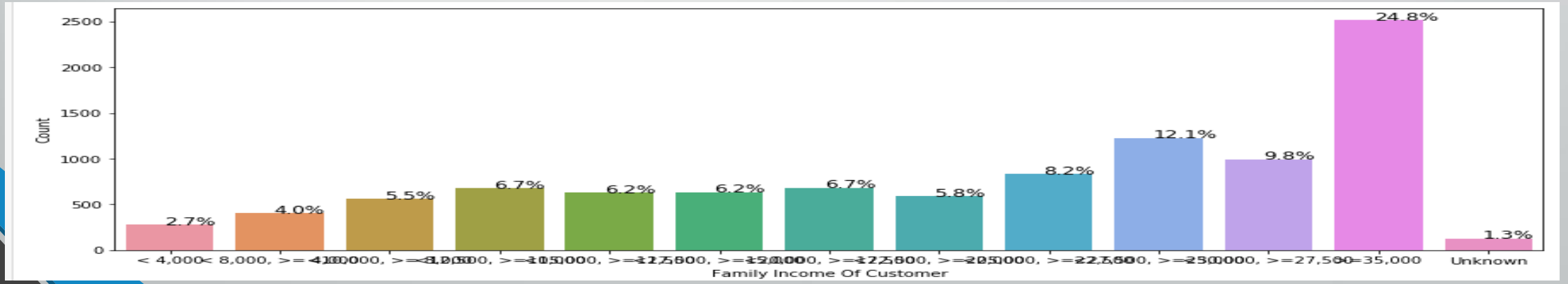
- Nearly 93% ; 9400+ customers' having their Own Home. Very interesting and strong customers to pull them into trading.



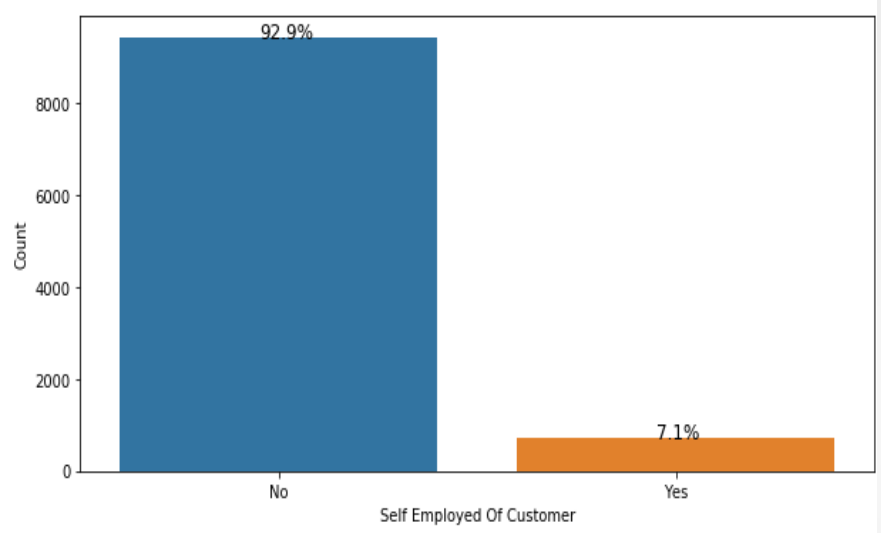
Family Income:

- Nearly 25% ; 2500+ customers having family income above $35,000 monthly.

- These customers are matured and financially well to do for Derivative market.

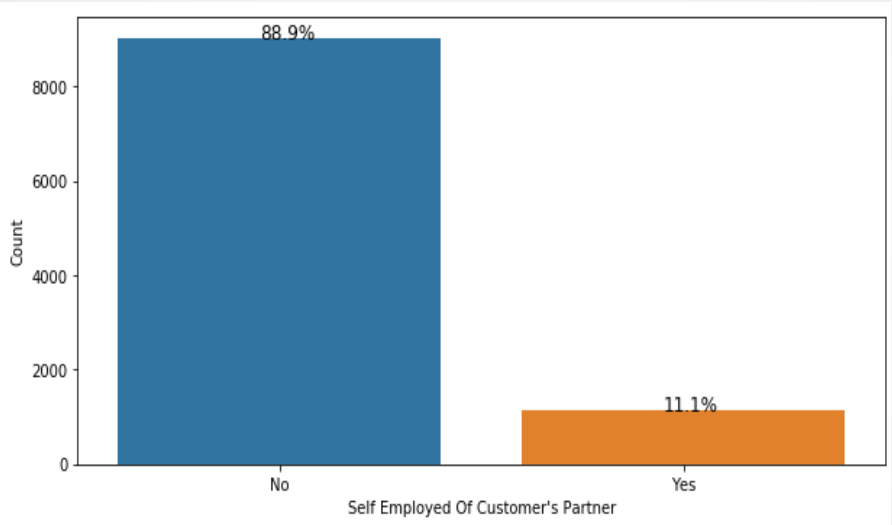


EDA – Self Employed & Partner's Status



Self-Employed Status:

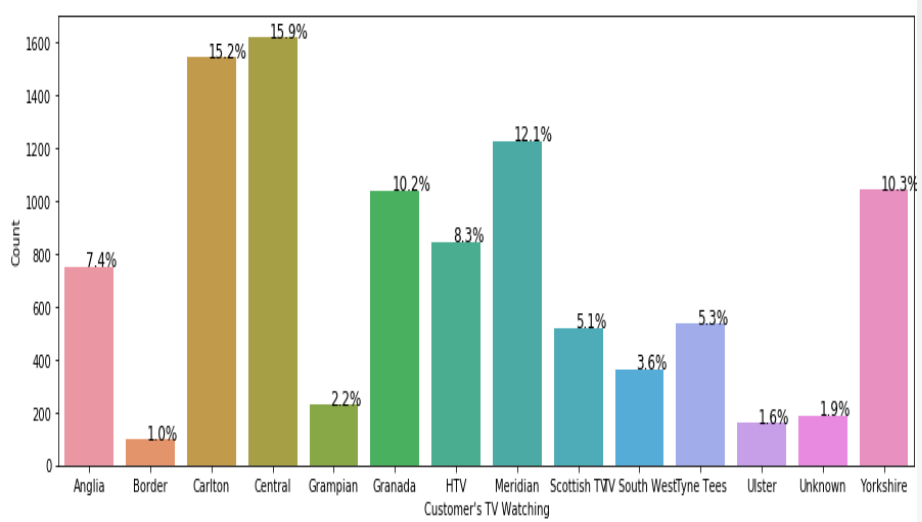
- 93% ; 9436 customers are not Self Employed; that means they are having regular income (salary based).
- These customer can take more Risk considering they have Fixed Income already.

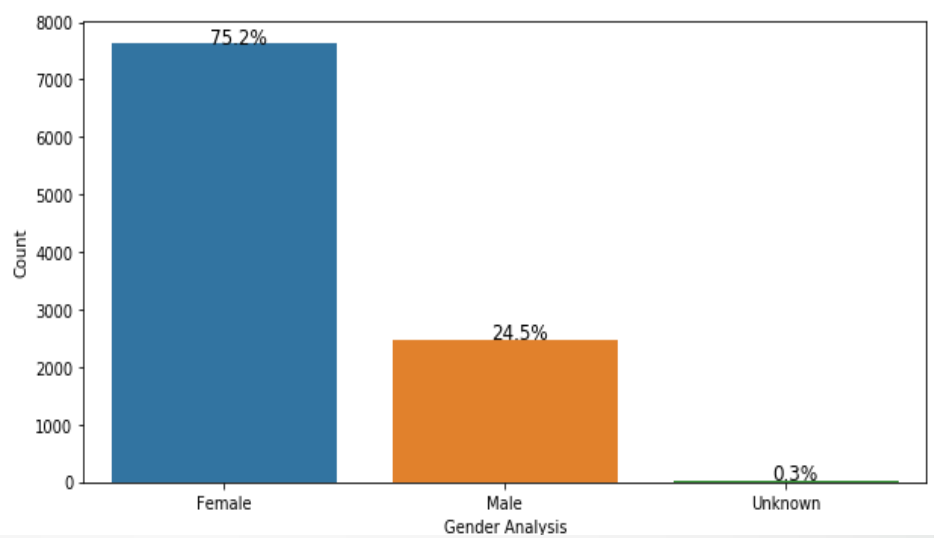


Partner's Self-Employed Status:

- 89% ; 9000+ customers partners are not Self Employed; that means they are having regular income (salary based).
- These customer along with their Partner's can take more Risk considering they have Fixed Income already.

EDA – TV Area & Gender





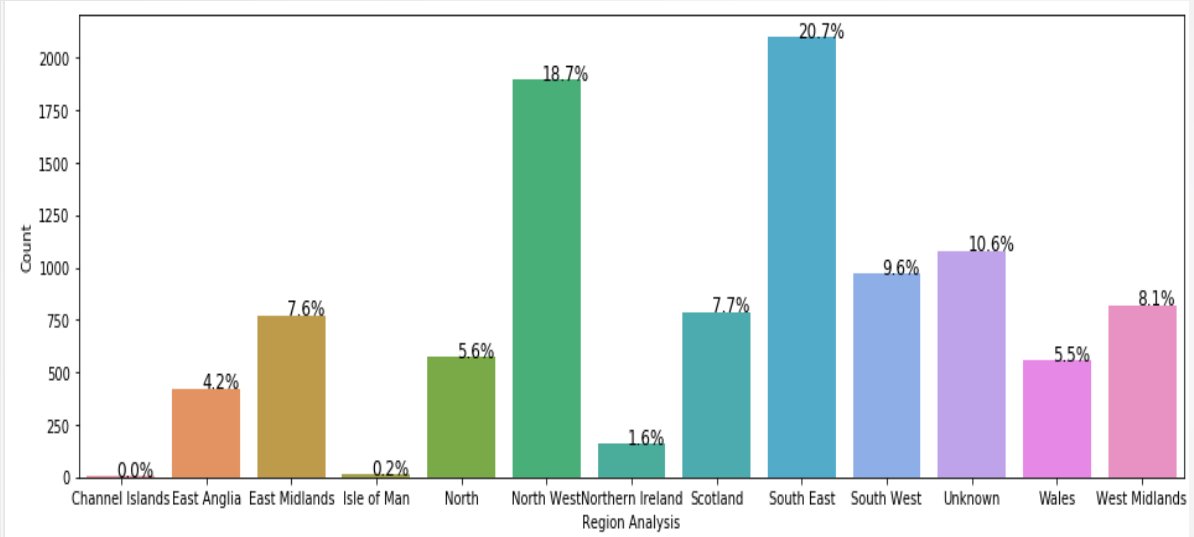
TV Watching:

- TV Channel Central and Carlton are mostly watched Channels, respectively 15.9% and 15.2%.
- Followed by 12.1% watch of Meridian.
- In total 13 TV Channels are watched; nearly 2% customer's channel watch time / preference not available/Unknown.

Gender:

- Most of the customer's are Female, 75.2%, 7634.
- Only 24.5% customers are Male.
- Scope to identify 0.3% customer's whose gender is Unknown.

EDA – Region



Region:

- There are 12 different region where Customer base is distributed with almost 11% Unknown regions as well.
- Nearly 21% ; 2100 Customer's are in Southeast region.
- Followed by Northwest with 19% Customer base.
- All others are below 10; Channel Island is 0% and Isle of Man is 0.2% (lowest customer base).

Bivariate Analysis Financial Institute

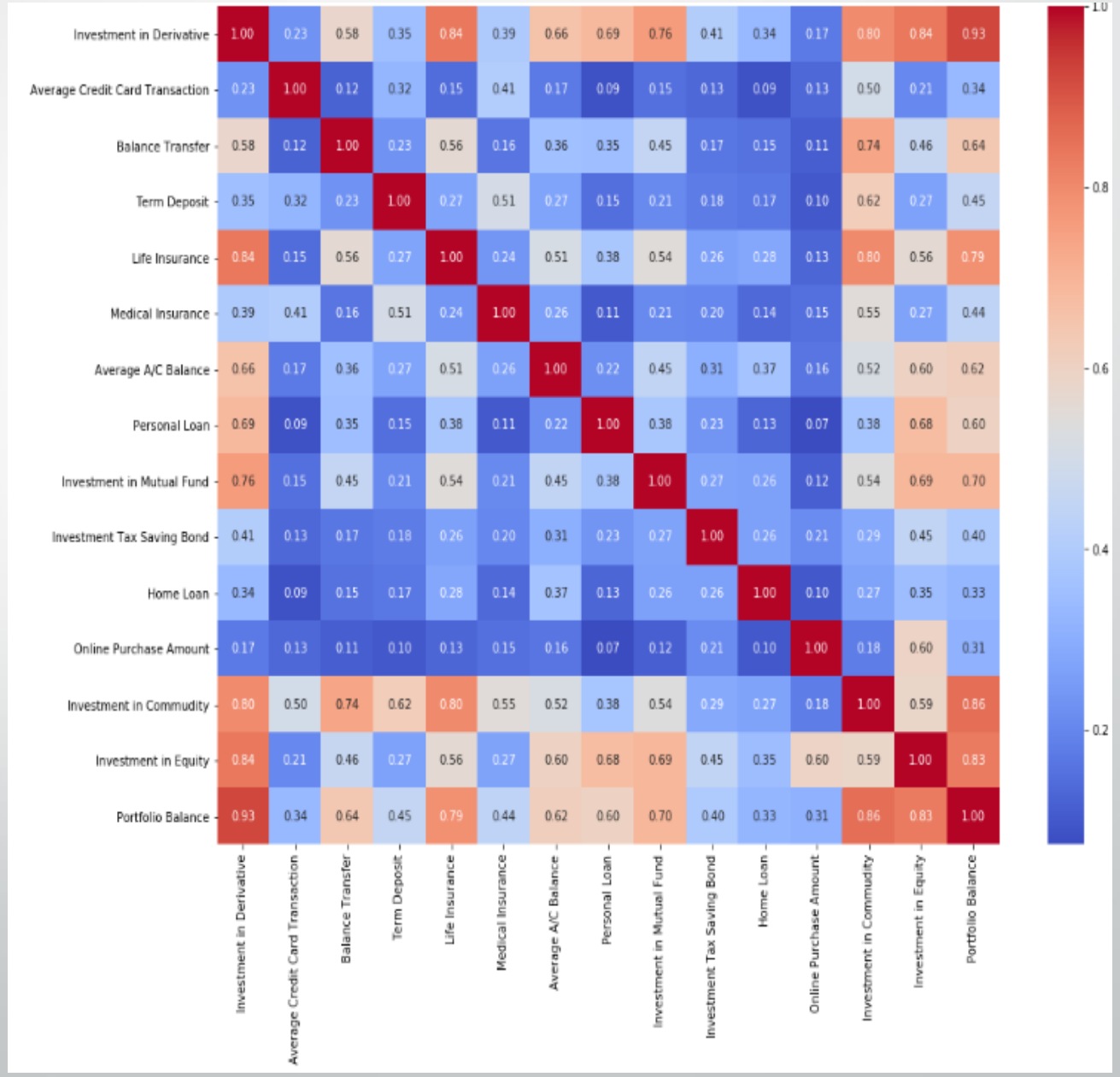
Bivariate Analysis - Heatmap

- If variable has correlation, it shows value 1
- If correlation value is nearby/towards 1, it shows positive correlation. That means increase in variable 1 impacts increase in variable 2.
- If correlation value is 0, that means there is no relation at all.
- If correlation value is nearby/towards -1, it shows –ve correlation. That means 1 variable increase and other decrease.

- Correlation Near/Towards +1
  - Portfolio Balance vs Investment in Derivative
  - Portfolio Balance vs Investment in Commodity
  - Portfolio Balance vs Investment in Equity
  - Investment in Equity vs Derivative
  - Investment in Commodity vs Derivative
  - Life Insurance vs Investment in Derivative
  - Life Insurance vs Portfolio Balance

- Correlation ZERO
  - Average Credit Card Transaction, Online Purchase Amount, Home loan are mostly linear, with all other numerical variables.
  - Average Credit Card Transaction, Balance Transfer, Term Deposit,
  - Medical Insurance vs Personal Loan, Home Loan, Tax Saving Bond treats mostly linear

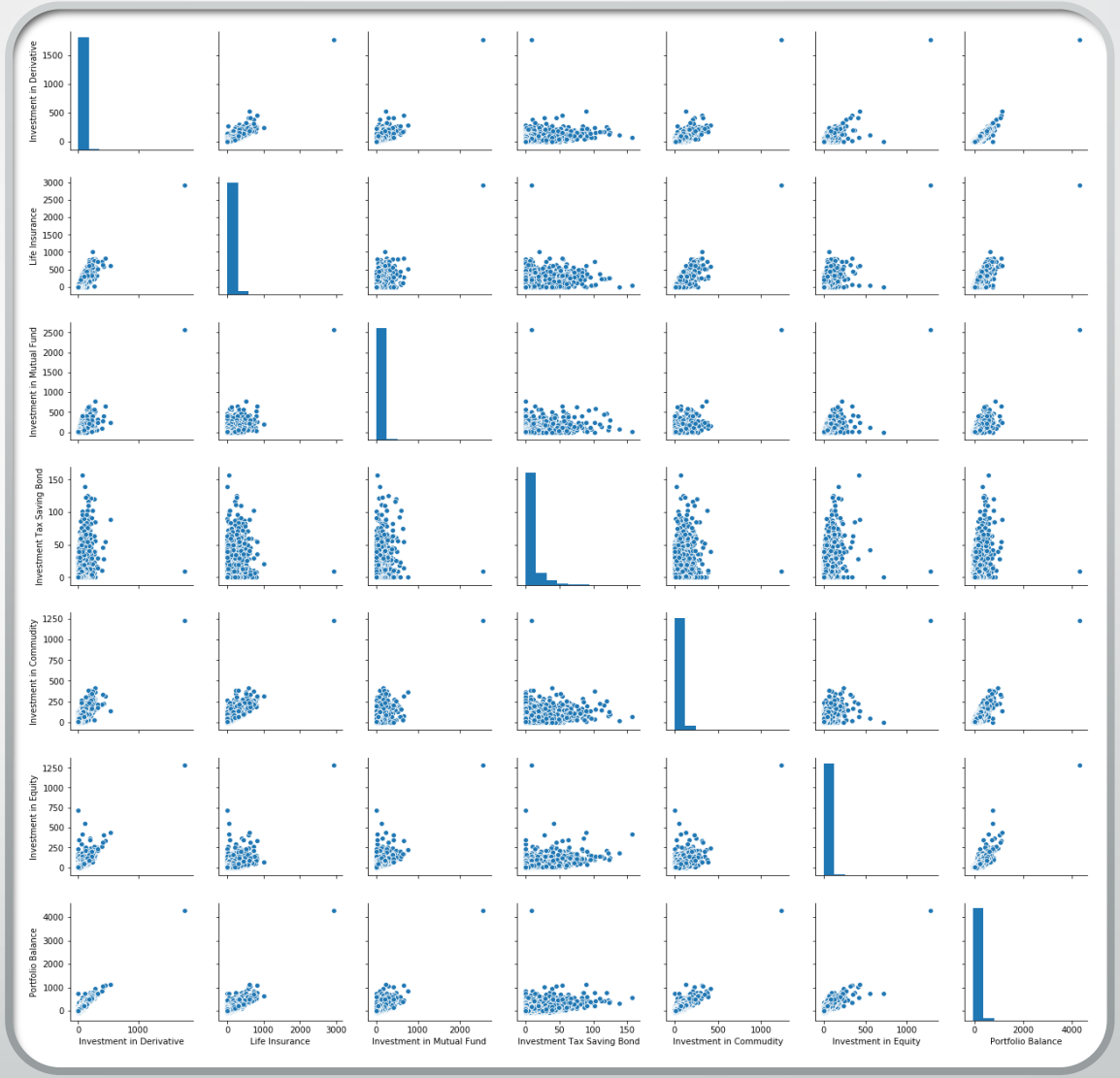
- We don't find –Ve correlation in this case.



Bivariate Analysis - Pairplot

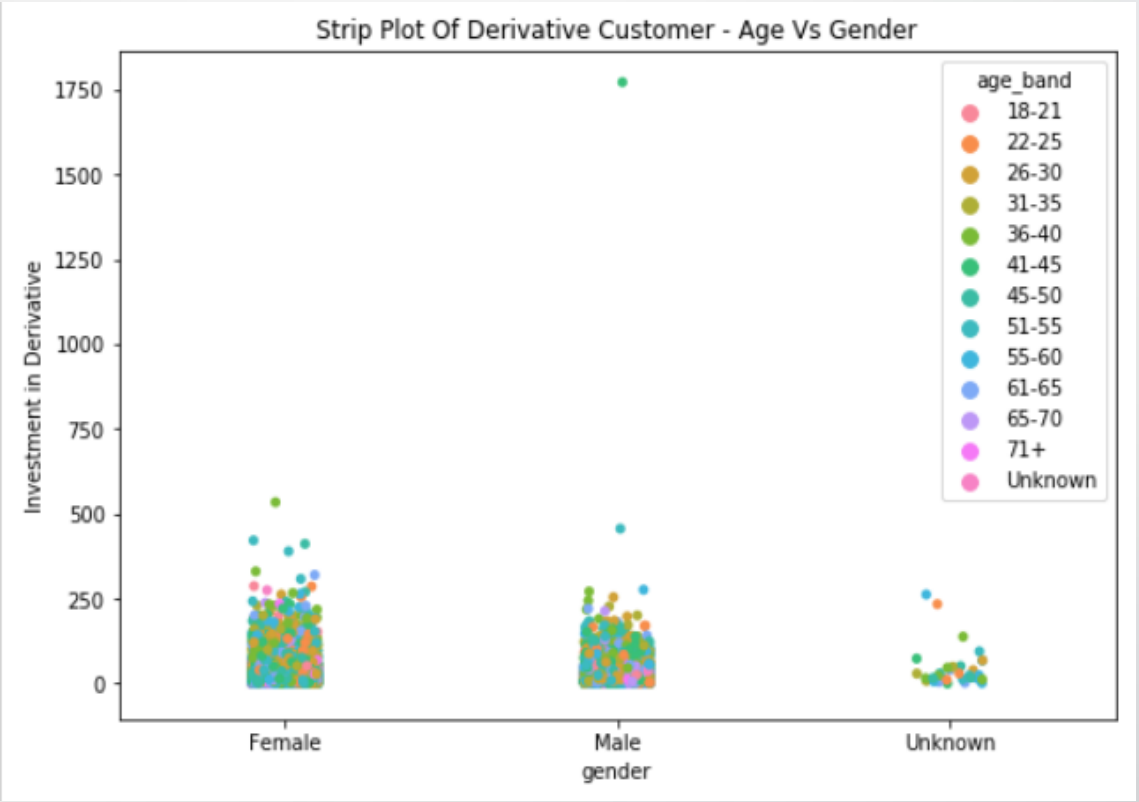
- There is no significant pattern observed in pairplot of below listed numerical variables.
  - Investment in Derivative,
  - Life Insurance
  - Investment in Mutual Fund
  - Investment Tax Saving Bond
  - Investment in Commodity
  - Investment in Equity
  - Portfolio Balance

- Means, at high level, Derivative/Trading doesn't have strong correlation with any other.

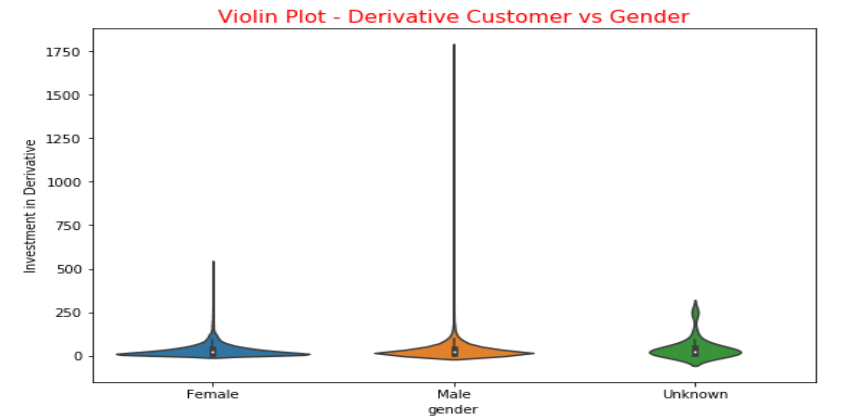


Bivariate Analysis – Strip Plot

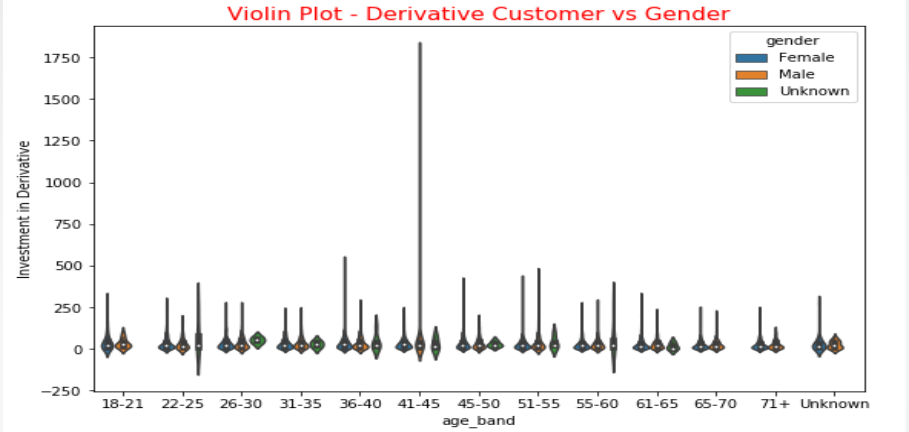
- Strip plot of Investment in Derivative vs Gender & Age Band clearly shows that
  - Higher number of FEMALE are involved into DERIVATIVEs (Trading) compared to MALE
  - Mostly 41-45, 46-50 and 51-55 are involved heavily in Derivatives along with few potential from 71+ and Unknown age group.



Bivariate Analysis – Violin Plots

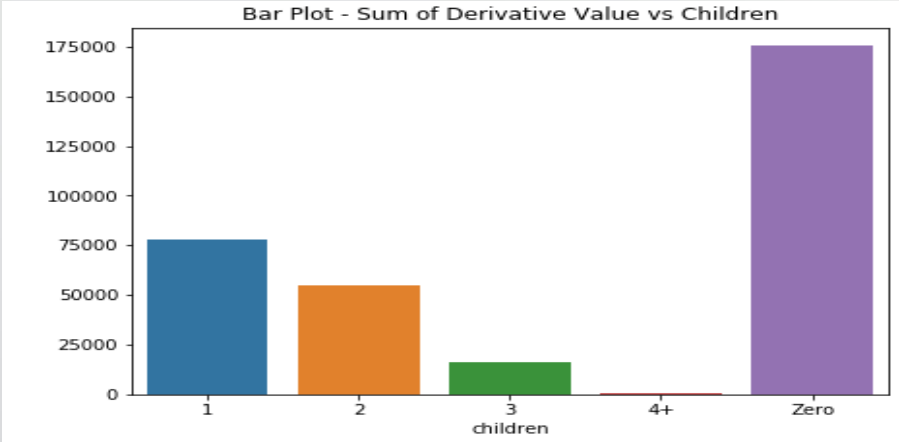


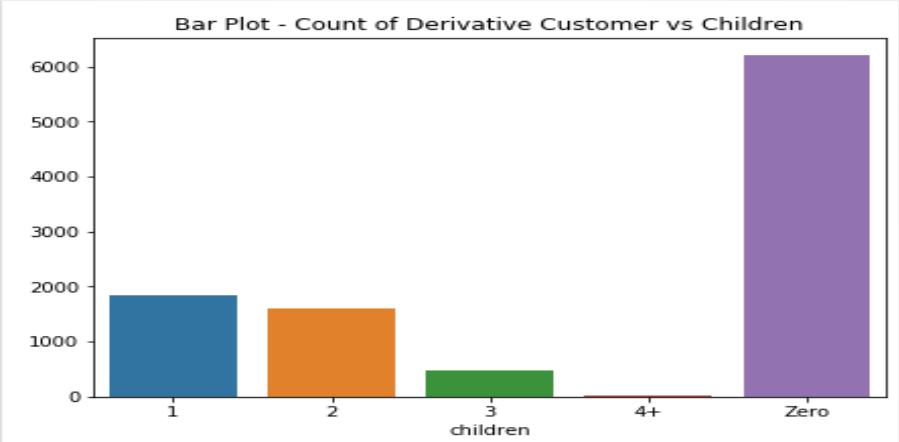
- Violin plot of Investment in Derivative vs Gender clearly shows that
  - FEMALE are more ACTIVE in derivative then MALEs.



- Violin plot of Investment in Derivative vs Gender and Age Band clearly shows that
  - FEMALEs and 41-45,46-50,51-55 age groups are more ACTIVE.

Bivariate Analysis – Bar Plots



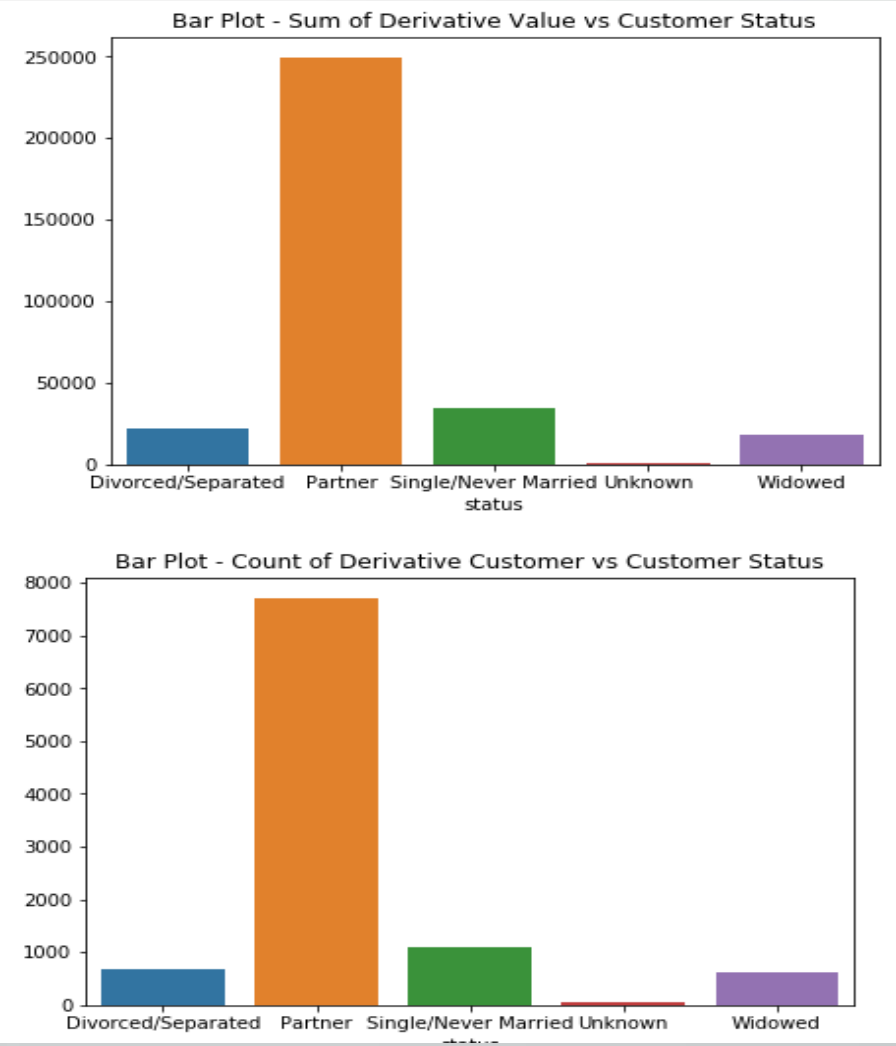


- Bar plot of Sum of Investment in Derivative vs Children clearly shows that
  - Customers with ZERO children are highly active Derivative players, nearly $175K.
  - Followed by 1 Child, nearly $75K.

- Bar plot of Customer Counts of Investment in Derivative vs Children clearly shows that
  - Customers with Zero Children, 6000+ are active Derivative players.
  - Followed by 1 Children Customers, nearly 2000.
  - Customers with 4 or more children don't participate in Derivatives (or negligible).

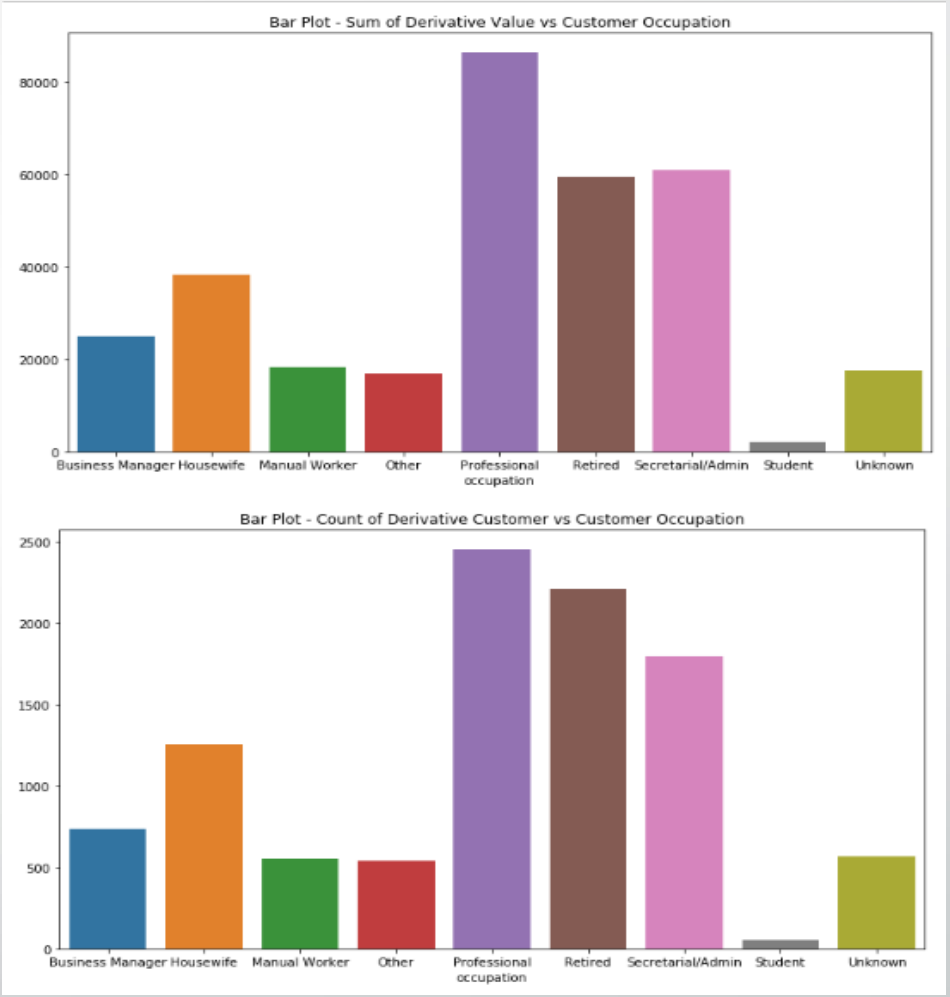
Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Status and Count of Investment in Derivative vs Customer's Status clearly shows that
  - Customers in Partner status are highly active Derivative players, nearly $250K.
  - Followed by Single/Never Married with close to $50K.
  - Nearly 8000 Customer with Partner status and 1000 with Single/Never Married Status are highly active.



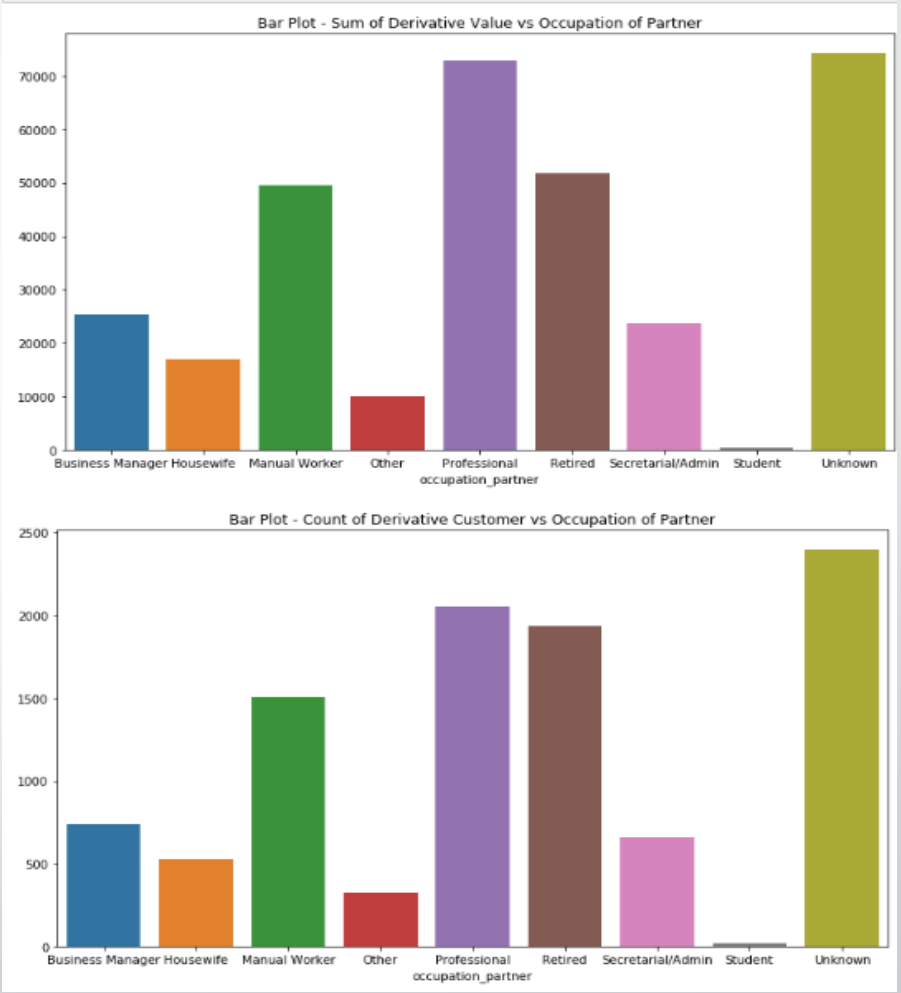
Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Occupation and Count of Investment in Derivative vs Customer's Occupation clearly shows that
  - Mostly Professionals (nearly 2400+) are doing heavy trading (derivatives) (nearly $80K)
  - Followed by Retired (nearly 2200) are doing Derivative of nearly $60K
  - Customers doing Secretarial/Admin job (nearly 1880) are also doing $60K Derivative turnover.
  - Housewives are also playing big role in Derivatives.
  - There are more than 500 Customers occupation is Unknown who are doing Derivative Turnover of nearly $20K.
  - Very rare/few Students does Derivative.



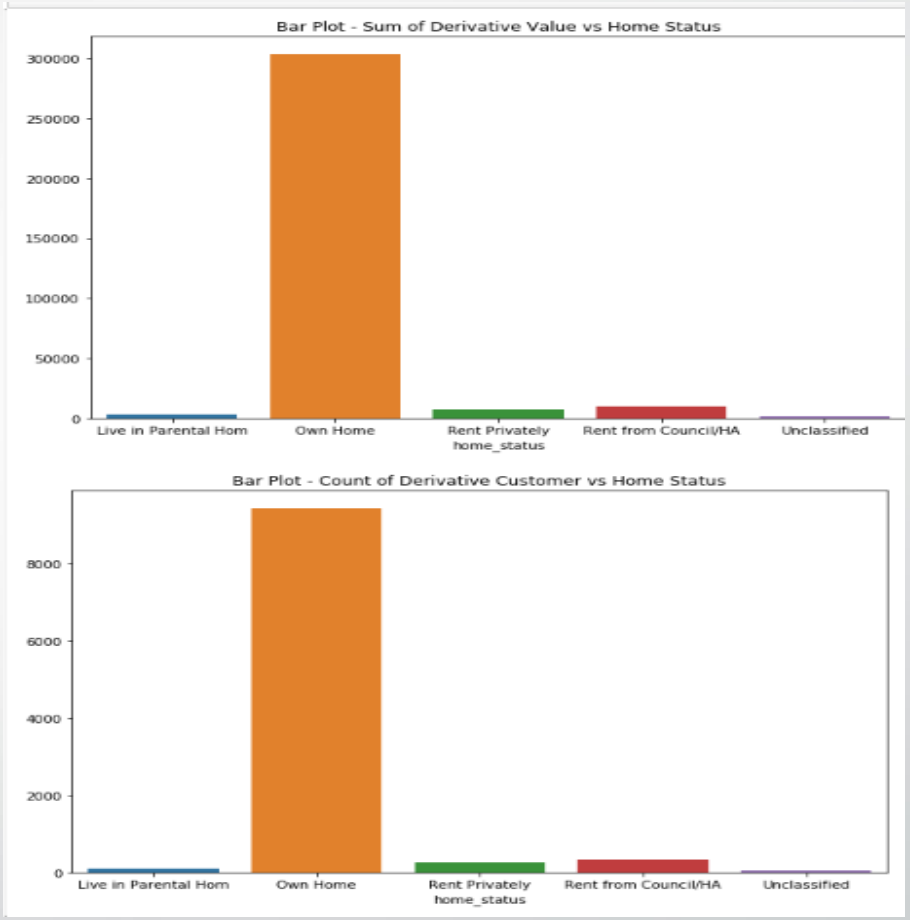
Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Partner's Occupation and Count of Investment in Derivative vs Customer's Partner's Occupation clearly shows that
  - Once again, Customer's whose partner is Professionals (nearly 2000+) are doing heavy trading (derivatives) (nearly $70K)
  - Followed by those whose Partner's occupation is Unknown (nearly 2400) are doing Derivative of nearly $70K
  - Customers with retired Partners are following them with nearly 1800 customer count and almost $50K Derivative Turnover.



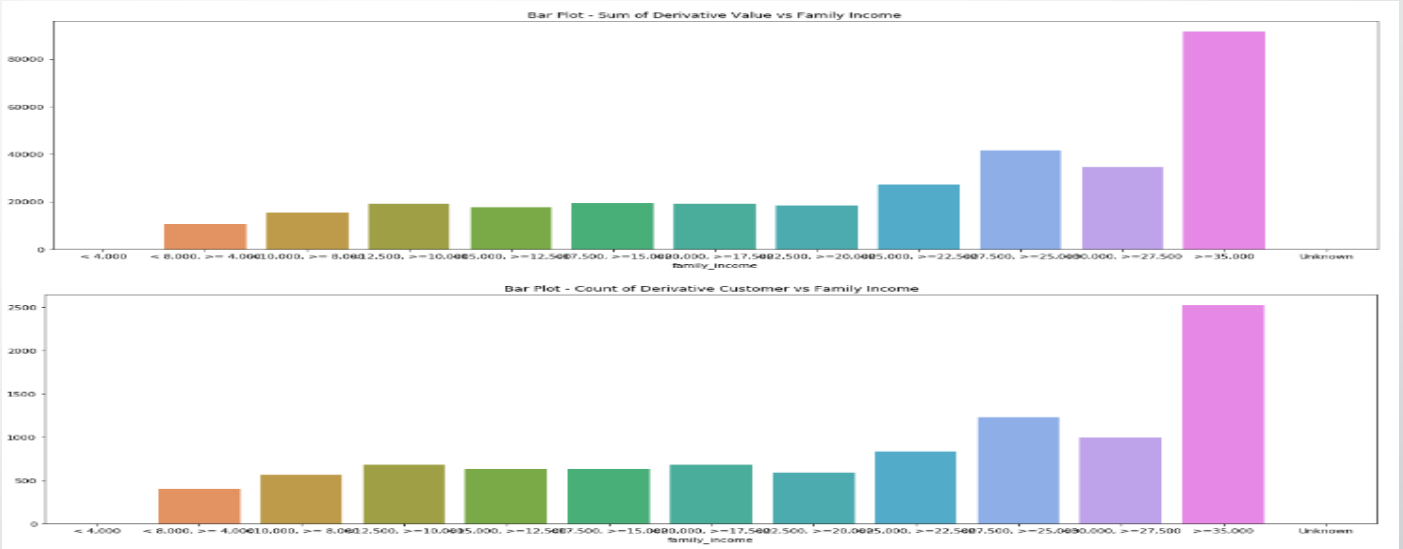
Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Home Status and Count of Investment in Derivative vs Customer's Home Status clearly shows that
  - Customer's owning Home, nearly 9000 are doing heavy derivatives turnover, nearly $300K



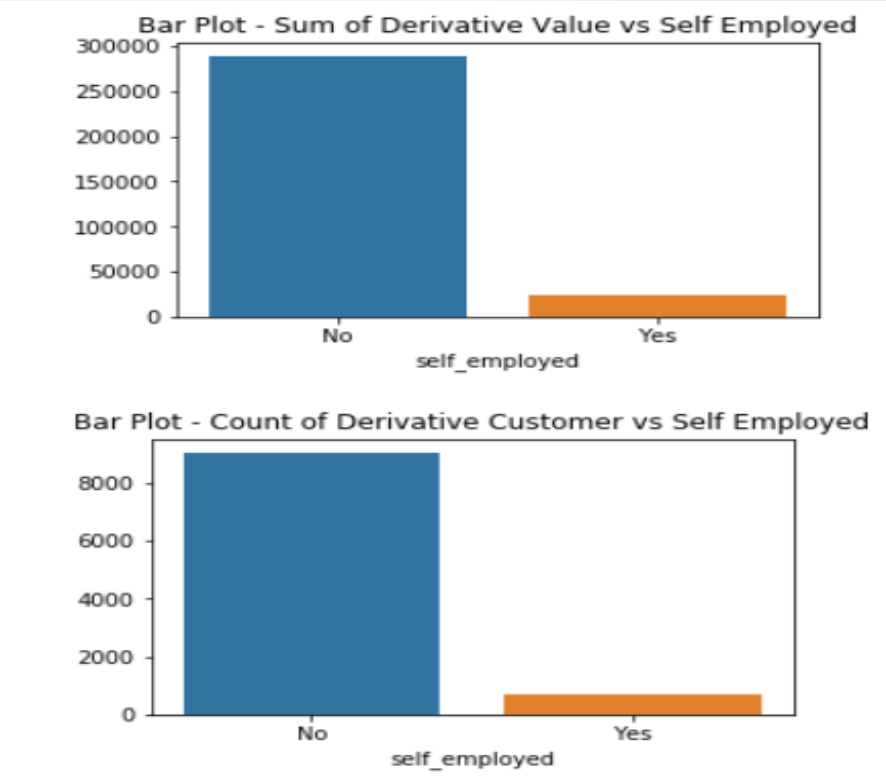
Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Family Income and Count of Investment in Derivative vs Customer's Family Income clearly shows that
  - Customer's having family income more than $35K are biggest contributor for Derivative Turnover, nearly 2500 Customers of this range income are doing heavy derivatives turnover, nearly $80K



Bivariate Analysis – Bar Plots

- Bar plot of Sum of Investment in Derivative vs Customer's Self Employed, and Count of Investment in Derivative vs Customer's Self Employed, clearly shows that
  - Customer's who are not Self Employed contributing heavily in Derivatives like nearly 9000+ customers contributing nearly $300K.



Multivariate Analysis Financial Institute

Multivariate Analysis

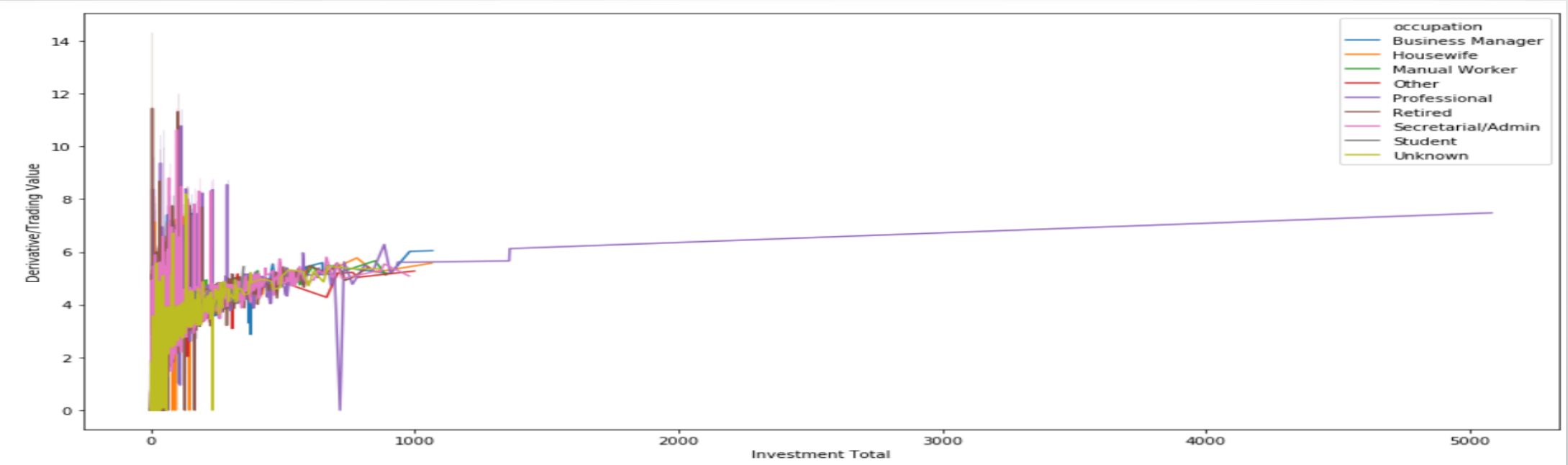
- Multivariate Line plot of
  - Investment Total* vs Sum of Investment In Derivatives basis Gender
    - Shows that MALEs are more focused on Investments and less on Derivatives
    - Whereas FEMALEs are more involved in Derivatives and less investments

*Investments : Equity, Commodity, Tax Saving Bonds and Mutual Fund

Multivariate Analysis

- Multivariate Line plot of
  - Investment Total* vs Sum of Investment In Derivatives basis Occupation
    - It's observed that Professionals are only category involved in both Investments & Derivatives

*Investments : Equity, Commodity, Tax Saving Bonds and Mutual Fund



Conclusion & Recommendations

- We analyzed a dataset of nearly 10k customers of the Financial Institute.
- It covers up Investments, Expenses, Demographics of customers.
- Data is spread in multiple segments like based on Children, based on Occupation, based on Home Status, based on Marital Status, based on Investments, based on Expenses etc.
- Mainly we are interested in how much Customer do/can do **TRADING/Derivatives** which is generating more commissions.
    - At high level, Customers who are Professional(s) by Occupation; Female by Gender, having Zero Children are doing more Trading/Derivatives.
    - They are the right customers to Offer Discounts. However, there are other Opportunity area as well.

Conclusions

- Customers who are **PROFESSIONAL**s by occupation are mostly involved in Derivative/Trading.
- Customers who are **FEMALE**s by gender are mostly involved in Derivative/Trading.
- Customers who have **ZERO** Children are mostly involved in Derivative/Trading.
- Customers are spread very heavily in Age Bands, Children, Occupations, Family Income groups etc
- Based on data, Average Credit Card usage is very limited/less, which is good sign in terms customer's are sensitive about their spending.
- Customers have monthly quality expense like Term Insurance, Medical Insurance, Life Insurance, Home Loan EMI etc.
- Overall, Expenses do not have any impact on Derivatives as per correlation plots. So we should avoid focusing on them.
- Customers have good quality of doing Investments for future in Mutual Fund, Tax Saving Bonds, Commodity and Long-Term Equities.
- Nearly 60% Customers do not have children. That means their risk appetite is high considering their expense expected to be lower.
- Nearly 40% customer falls in 26 to 45 age group, which is again showing there are potential customers who has higher risk appetite.
- Nearly 75% customer marital status is living with their Partner (assumed spouse). These are matured couples/families who can do better fund management.
- Southeast & Northwest are biggest customer zone. 75% Females are Institution's customer as of now.
- There is positive correlation among Portfolio Balance - Investment In Commodity - Investment in Equity - Life Insurance vs Derivatives.

Recommendations

- There are 24+% Customer are Professional by occupation. That shows potential of heavy Trading/Derivative.
- Nearly 3% Customer's are owning home. That means either they are repaying home loan or paid already. That shows potential.
- Based on family income, great potential found as below :
  - Nearly 25% customer's with >35,000 family income;
  - Nearly 12% customer's with between 25000 to 27500
  - Nearly 10% customer's with between 27500 to 30000
- Nearly 90% customer's are not Self Employed; whereas 89% partner's are not Self Employed. There is potential.
- Female customer and middle age group are heavily involved in TRADING. Keep them motivated with relative offers.
- Customer's with No children are highly involved in TRADING followed by having 1 child. Keep them motivated with relative offers.
- Marital status as Partners are highly involved in TRADING. Keep them motivated with relative offers.

Future Analysis

- Further analysis can be done on finding potential customer doing/can do Derivative with base Total Investment vs Total Expense.
- There are nearly 21% Customer, and 20% Partners are Retired. They can be further investigated to find an opportunity.
- There are 5% Customer, and 24% partner's profession is Unknown. That can be an opportunity to identify their profession and analyze.
- Customer and Partner, both are Not Self Employed. These can be further analyzed to find an opportunity.
- Customer who are Not Professionals, further analysis can be done to find an opportunity.
- Customer of Male and Unknown category, further analysis can be done to find an opportunity.
- There is lot of scope of detailed analysis considering it's sizable database with 32 total variables and 10k+ observations..

Learnings

- **DISTRIBUTION:**
  - There are 3 types of Distributions.
  - UNIVARIATE
    - Focus is only on ONE variable. It's simplest analysis form.
    - Major purpose of UNIVARIATE is to find pattern or trend in that variable (in Database).
    - We can describe patterns using central tendency (mean, median and mode) and dispersion (range, variance, standard deviation, maximum and minimum values and interquartile range).
    - Histogram and Barcharts are widely used here.
  - BIVARIATE
    - Here we have focus on TWO variables.
    - In this type of analysis, we tend to find Cause & Relationship between them.
    - It includes Scatter Plot, Regression Analysis and finding correlation coefficients.
  - MULTIVARIATE
    - When we focus on THREE or MORE variables for analysis, it's Multivariate.
    - It's mainly used for more complex sets of data for large datasets.
    - Heatmaps, Scatter Plots are widely used here.