# Austo Project

Python Project

Submitted on 22nd of October, 2024

# Contents

# Business Problem Overview and Solution Approach

**The Problem**

- Cars have become important means of transportation in today's world. People purchase cars for different reasons related to people's needs, preferences and necessities.
- Austo is an automobile company interested in growing its business in Europe.
- The company is interested in knowing the factors that affect the purchasing behaviour of their customers in US and the types of cars that customers prefer.
- The company is seeking data driven advice in order to meet the demands of their customers in US

**Solution Approach**

- The approach is to draw insight from the data provided, analyze the data by performing exploratory data analysis that can help help the company make informed decision needed to manipulate their business strategy and production thereby meeting the demands of their customers

# Data Background and Content

The data contains the different buyer's data corresponding to different types of products(cars). The data has 1581 rows and 14 columns. The columns types are integer, and object (6 int and 8 object). The data has the following columns:

- Age: Age of the customer
- Gender: Gender of the customer
- Profession: Indicates whether the customer is a salaried or business person
- Marital_status: Marital status of the customer
- Education: Refers to the highest level of education completed by the customer
- No_of_dependents: Number of dependents(partner/children/spouse) of the customer
- Personal_loan: Indicates whether the customer availed a personal loan or not
- House_loan: Indicates whether the customer availed house loan or not
- Partner_working: Indicates whether the customer's partner is working or not
- Salary: Annual Salary of the customer
- Partner_salary: Annual Salary of the customer's partner
- Total_salary: Annual household income (Salary + Partner_salary) of the customer's family
- Price: Price of the car
- Make: Car type (Hatchback/Sedan/SUV)

**Data Background and Content (continued)**

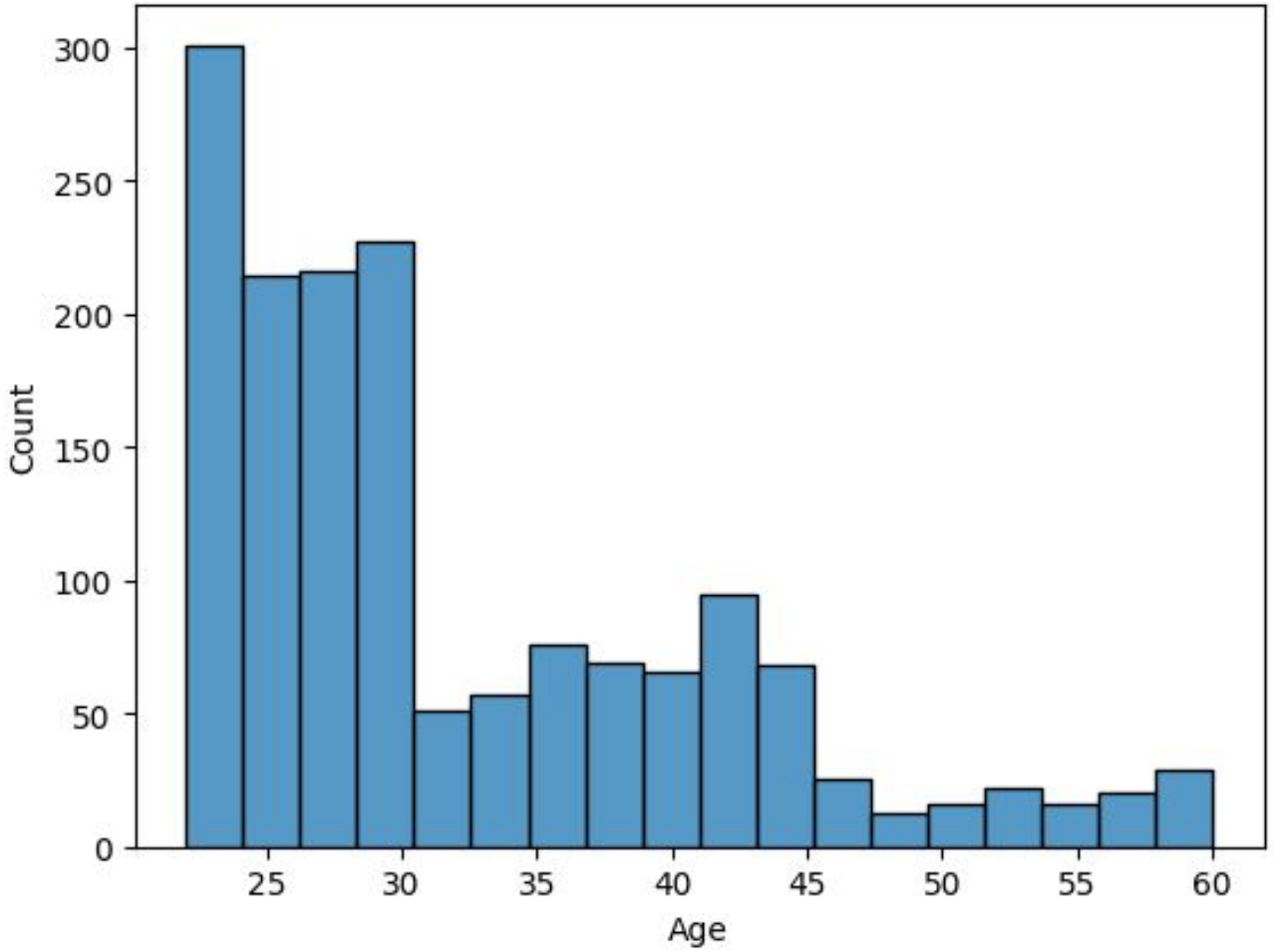
- The age of customers ranges from 25 to 60 years, their mean age is around 32.
- There are 1252 males customers.
- There are more customers who earn salaries compared to customers who run a business
- Most of the customers are married, 1443 customers are married
- There are 985 customers with postgraduate degrees
- The number of dependants of customers ranges from 0 to 4
- There are 792 customers with personal loan
- There are 1054 customers with house loans
- Most customers have partners who work

# Data Background and Content (continued)

- The salary of customers range from 30000 to 90000 with mean around 59732.44
- Customers have partners whose salaries range between 0 to 80000
- Customers total salary range from 61000 to 158000 Price of car range from 18000 to 70000
- There are three make/car types, customers buy Hatchback most
- There are no missing values in the data given
- There are 237 SUV cars

# EDA Results

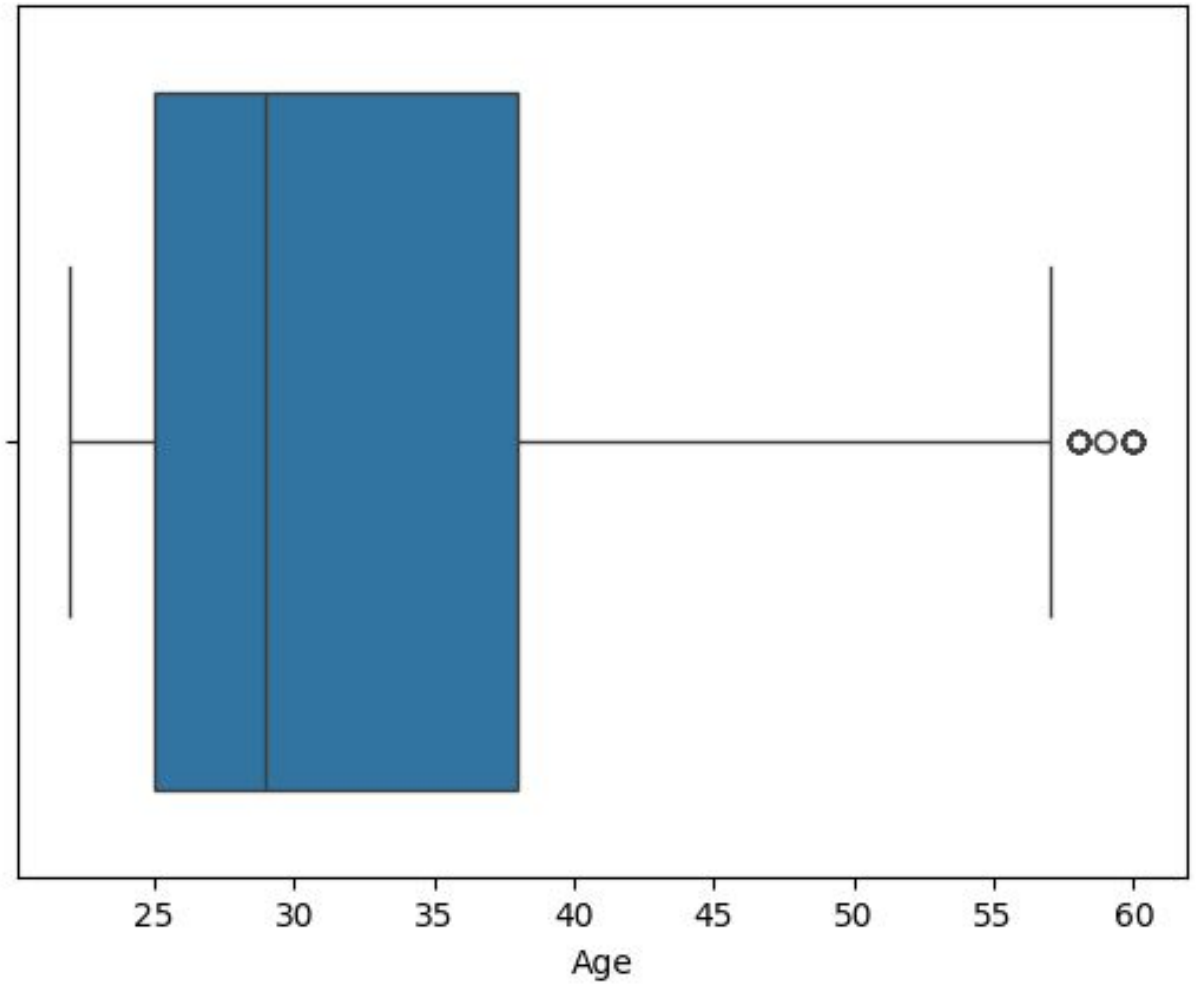
## Univariate analysis: Age




- Most customers who by cars tend to be around 30 years old or less
- People tend to buy less cars as they get old
- The age range 46-48 years old tend to have the least number of customers who buy cars

# EDA Results

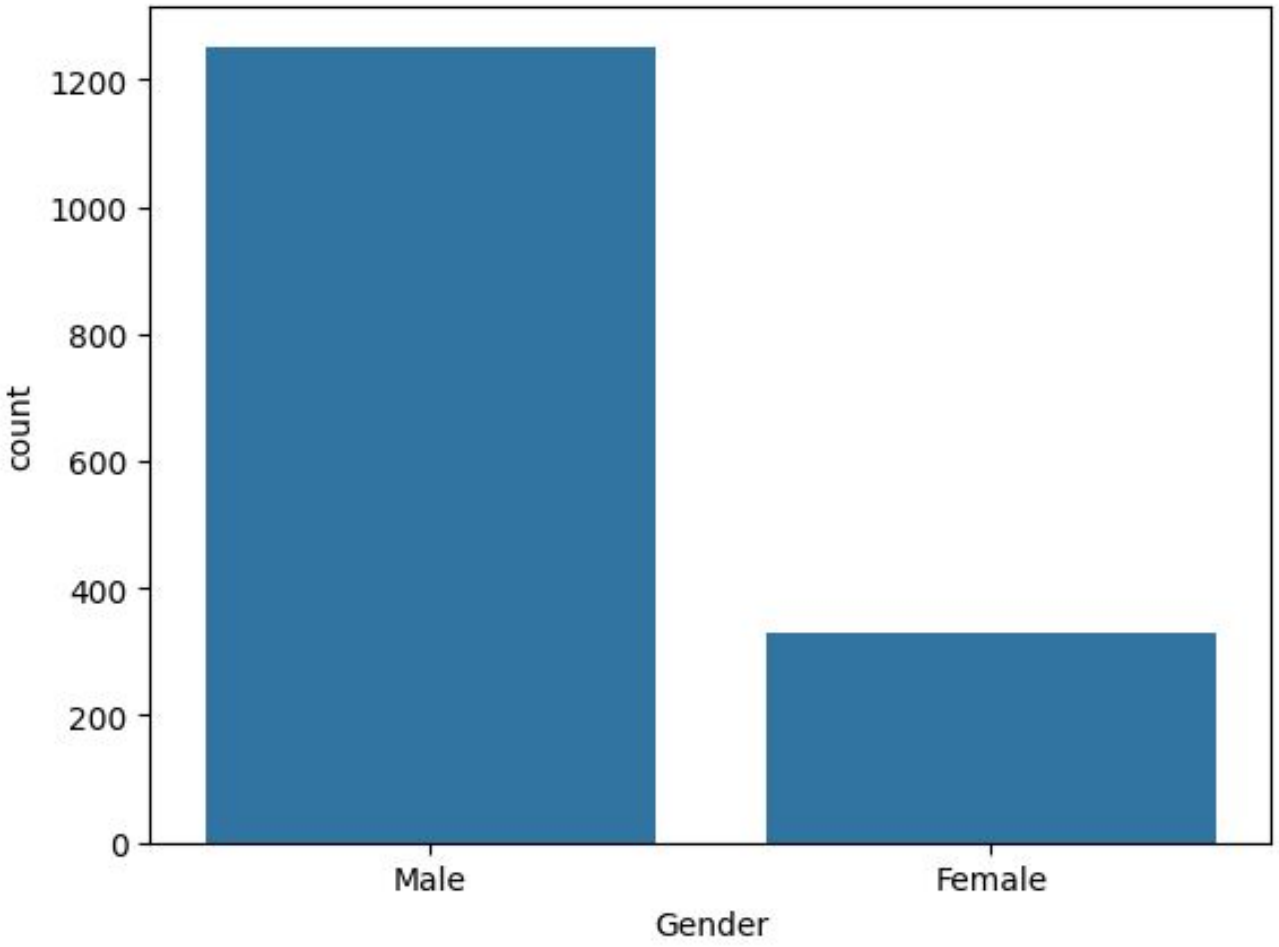
## Univariate analysis: Age




- There are outliers. This is evident from the boxplot. Some customers who are quite old still buy cars
- The median age of customers is around 29
- The distribution is rightly skewed

# EDA Results

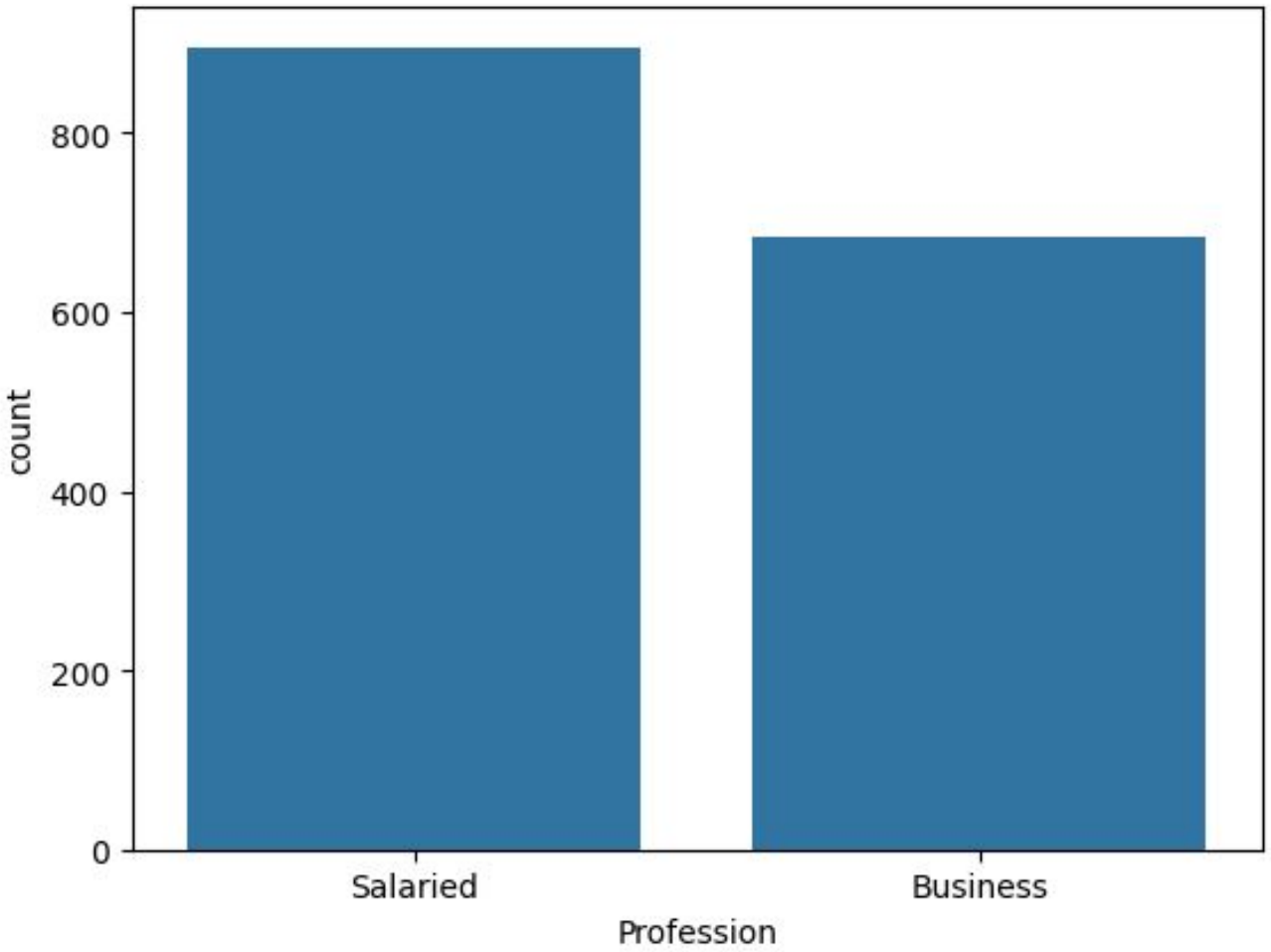
## Univariate analysis: Gender




- There are male and female unique gender types
- There are more males than females as shown in the plot
- Male customers tend to buy more cars than their female counterparts

# EDA Results

## Univariate analysis: Profession

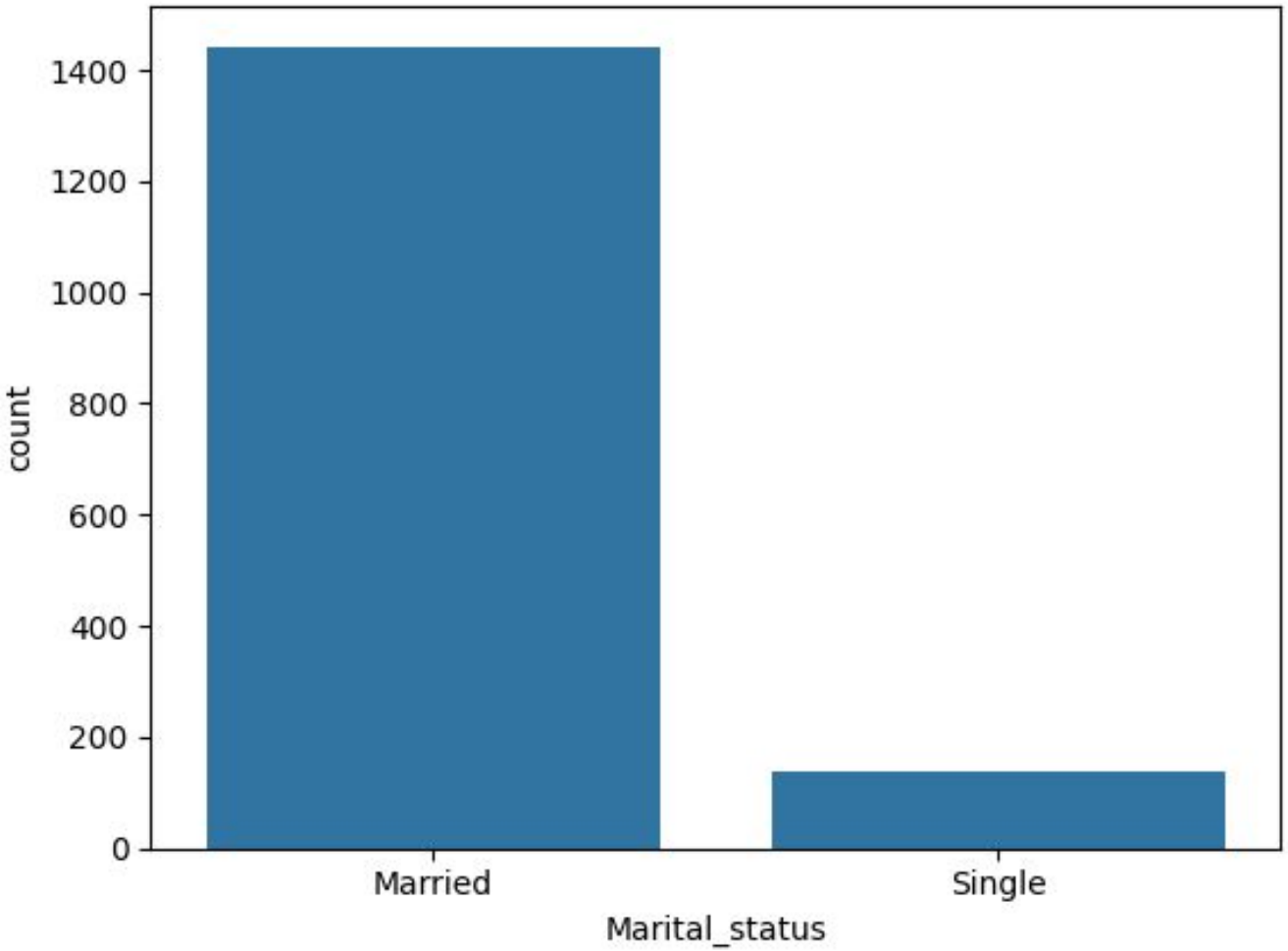



There are two profession types: salaried and business

There are more salary earners than business people who buy cars

# EDA Results

## Univariate analysis: Marital status

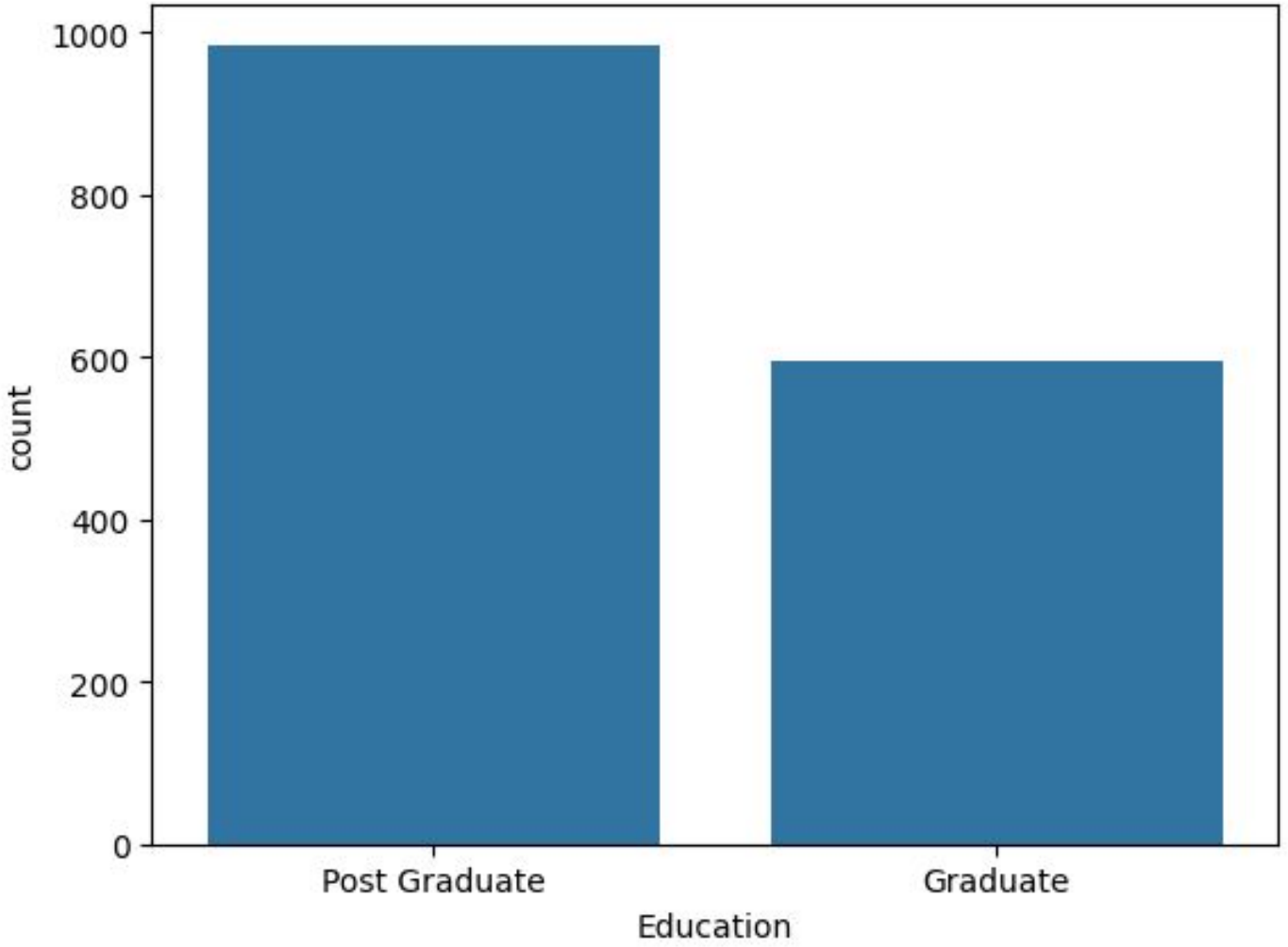



There are more married customers than single

1400 married customers buy cars

# EDA Results

## Univariate analysis: Education



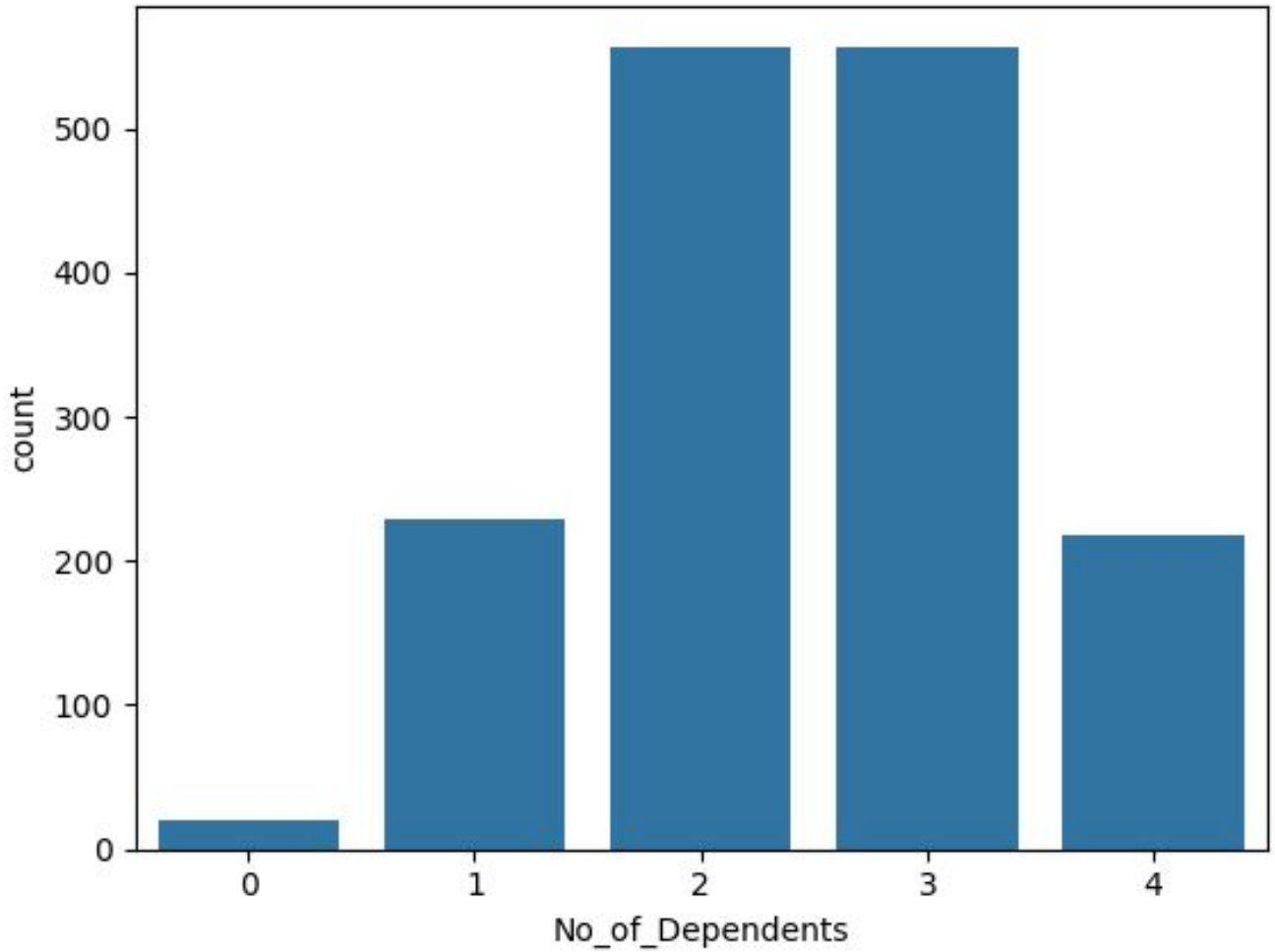

Customers have either a graduate degree or a postgraduate degree

There are more customers with postgraduate degrees

Less than 600 customers have a graduate degree

# EDA Results

## Univariate analysis: Number of dependents



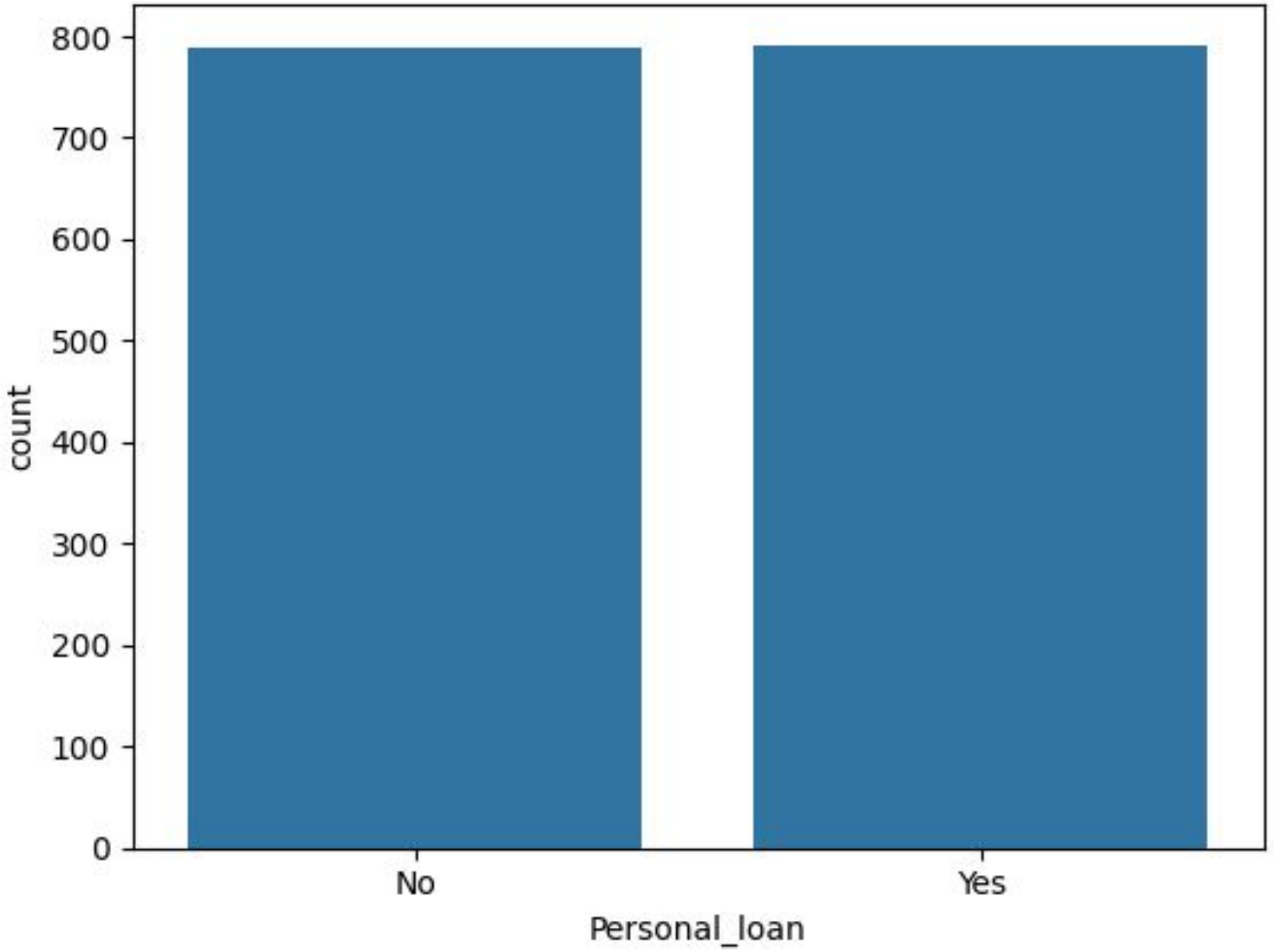

The customers have either 0, 1, 2,3 or 4 dependents

Most of the customers have either 2 or 3 dependents

There are very few customers with no dependents.

# EDA Results

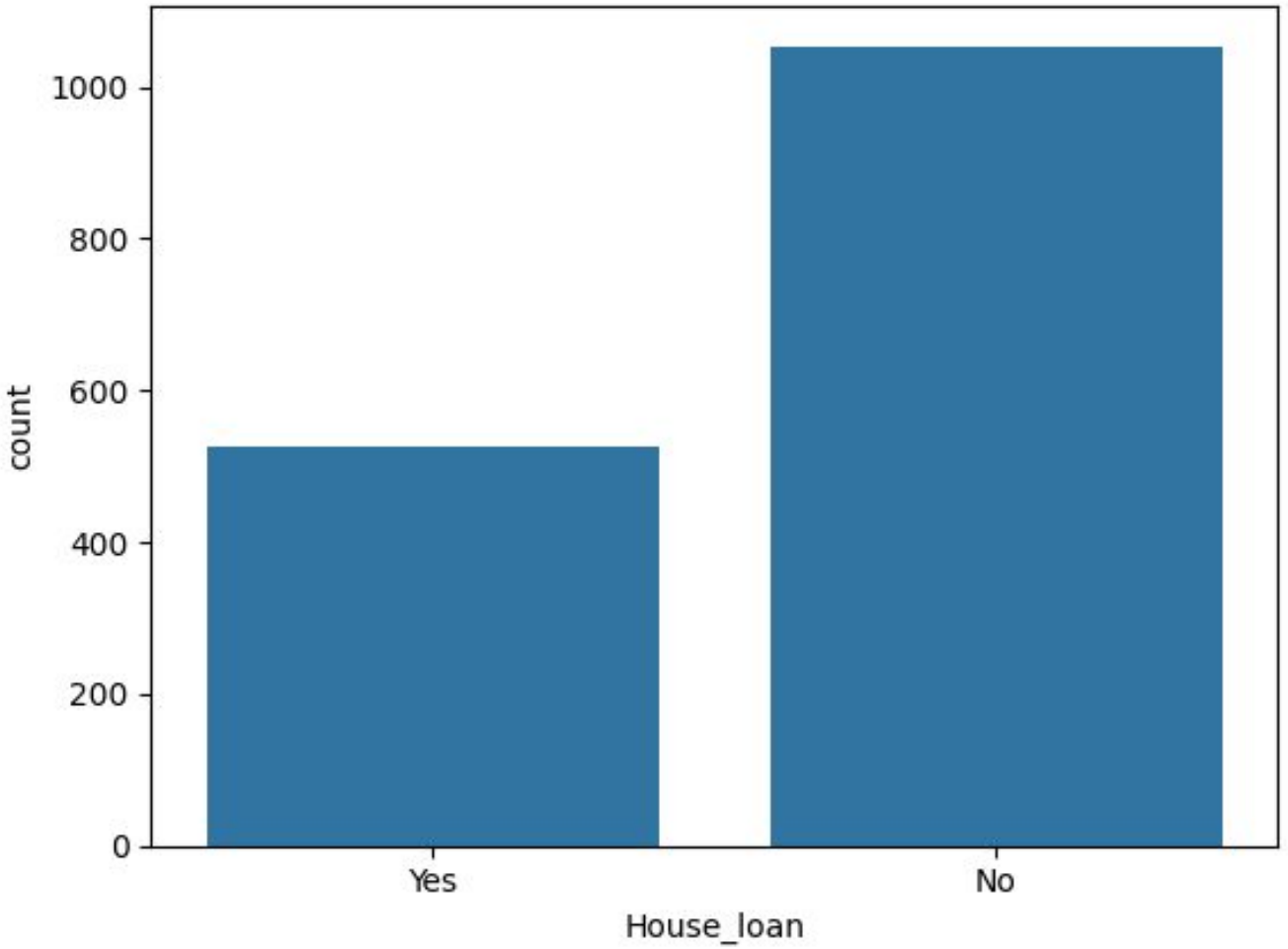
## Univariate analysis: Personal loan




- Customers either have  personal loan or not
- The number of customers who have personal loan and those who don’t seem to be the same

# EDA Results

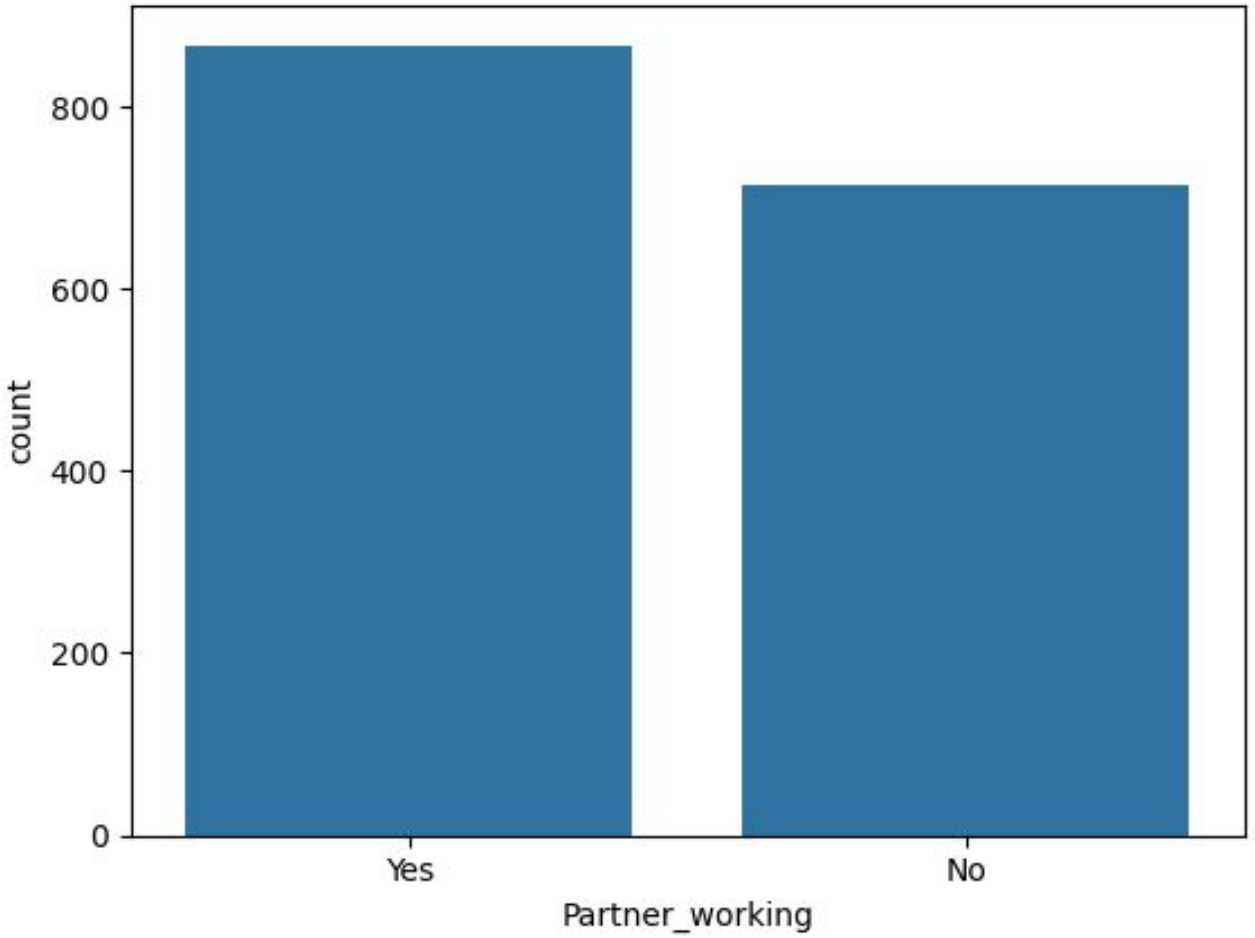
## Univariate analysis: House loan




- Customers either have house loan or not
- There are more customers without house loan

# EDA Results

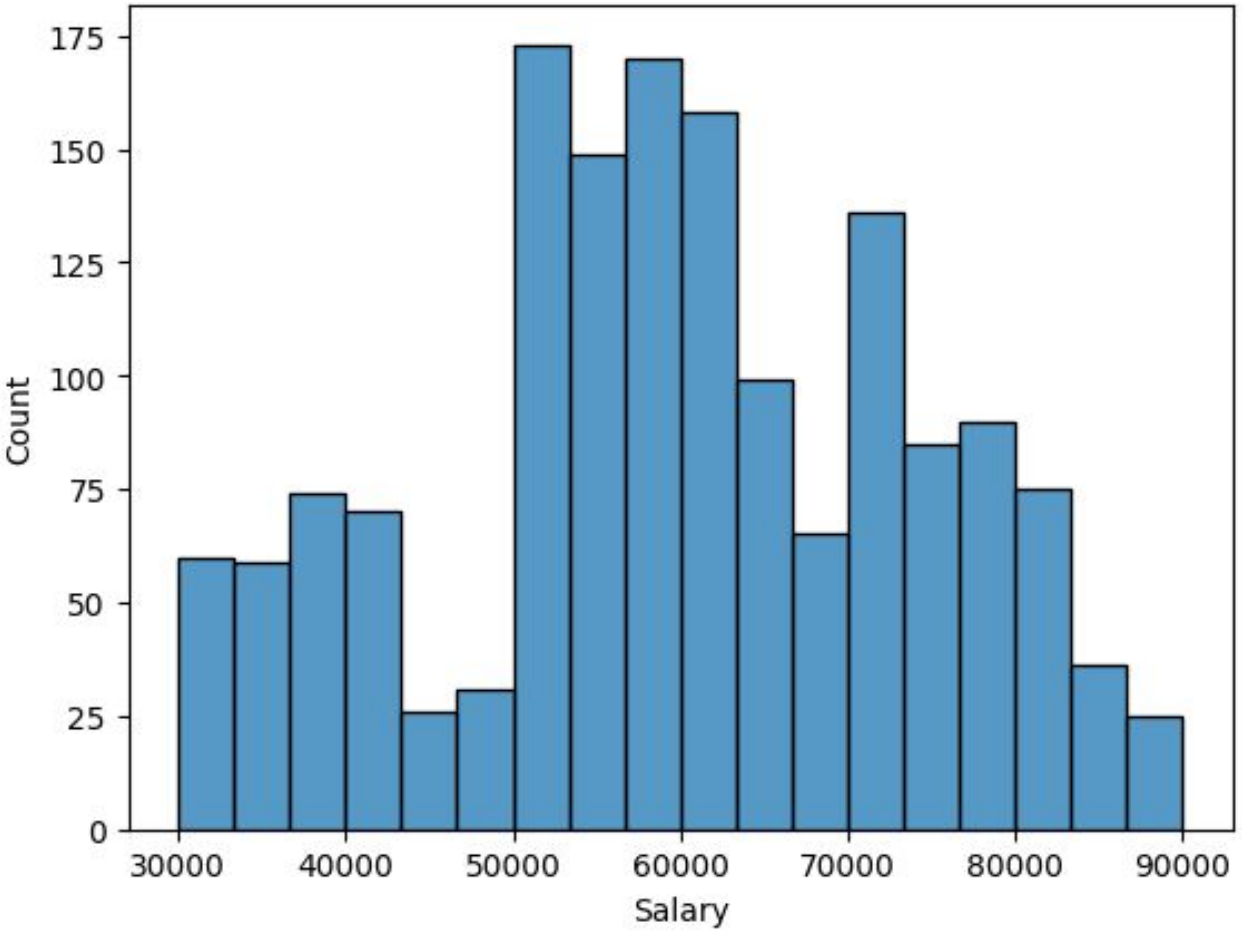
## Univariate analysis: Partner working




- Some customers have partners who are working, while others do not.
- There are more customers who partners work than customers whose partners do not.
- More than 800 customers have partners who work

# EDA Results

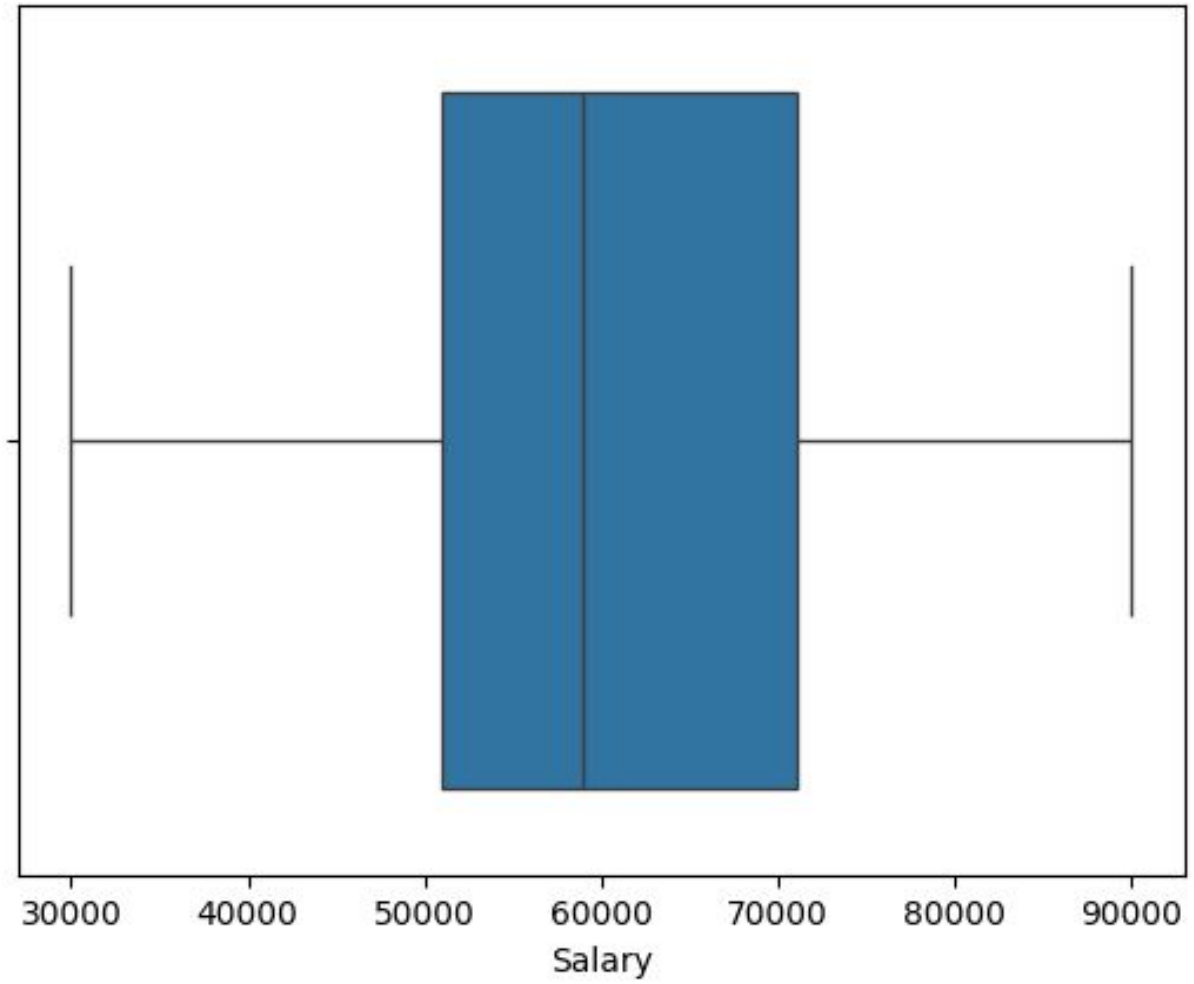
## Univariate analysis: Salary




- Distribution is not normally distributed.
- There are more Customers who earn between 50000 and 65000

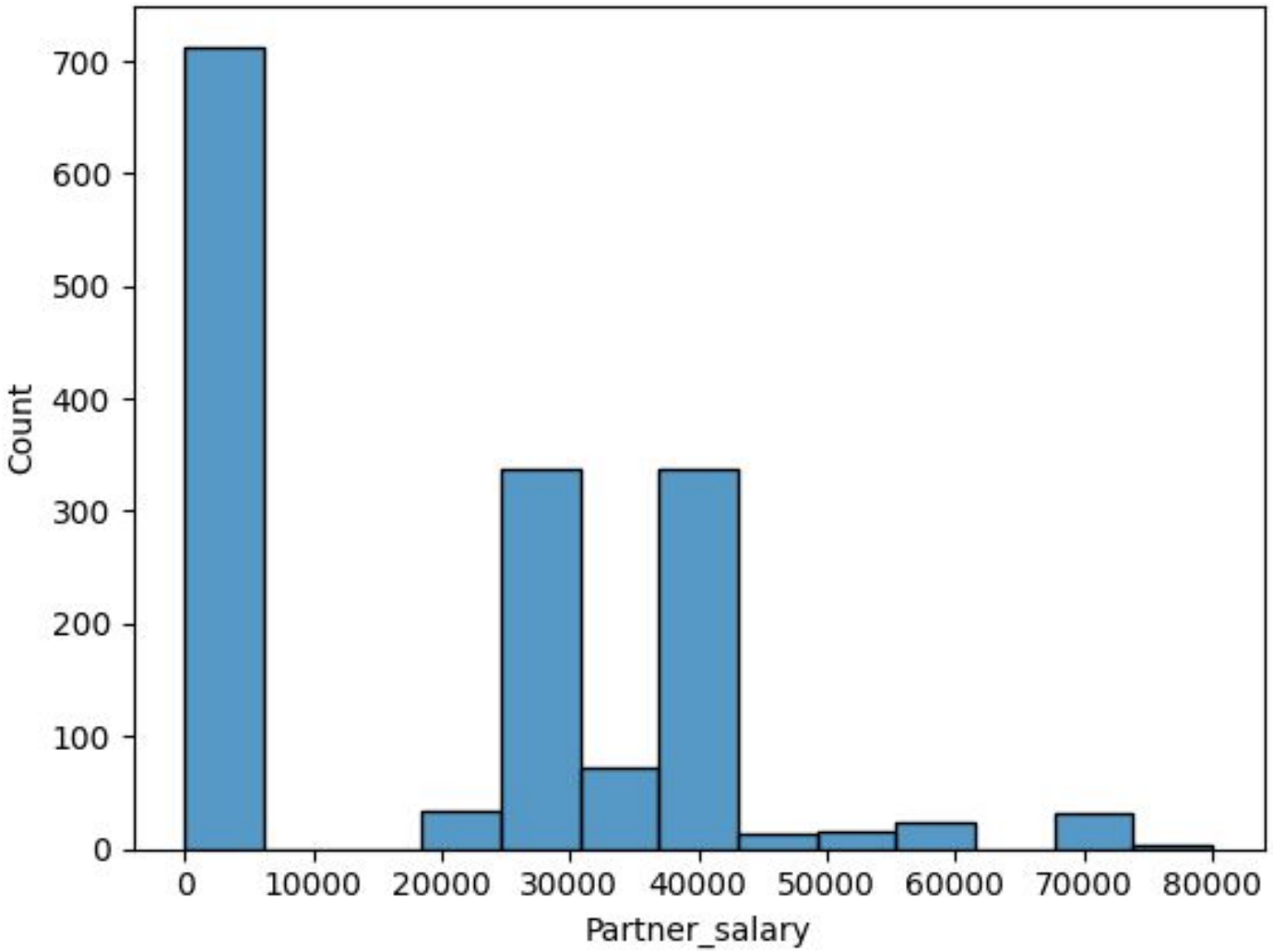
# EDA Results

## Univariate analysis: Salary (continued)




- The median salary of customers is around 59000
- There are no outliers

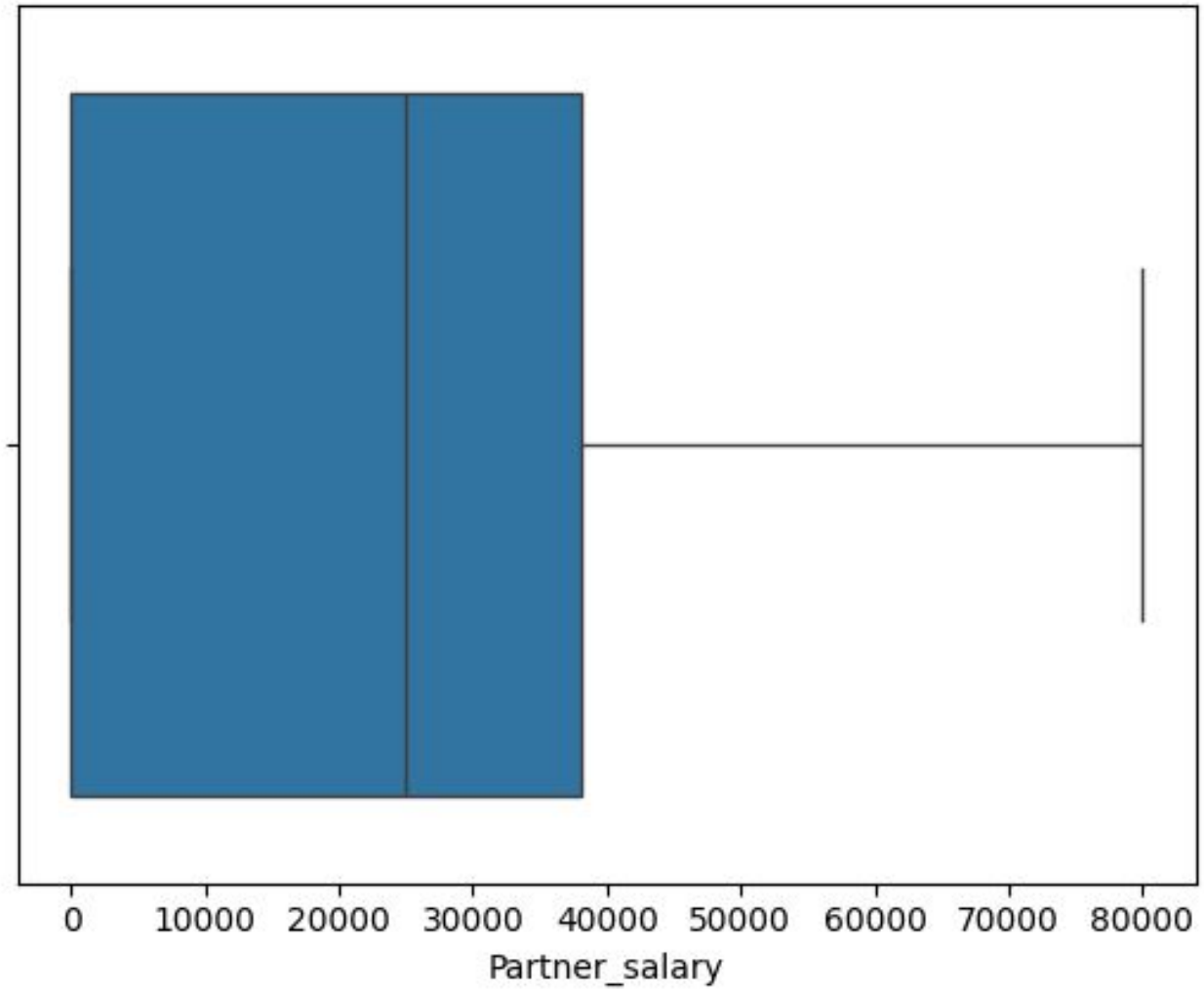
# EDA Results
# Univariate analysis: Partner's salary




- About 700 customers have partners who earn less than 10000
- There are relatively few customers whose partners earn more than 45000

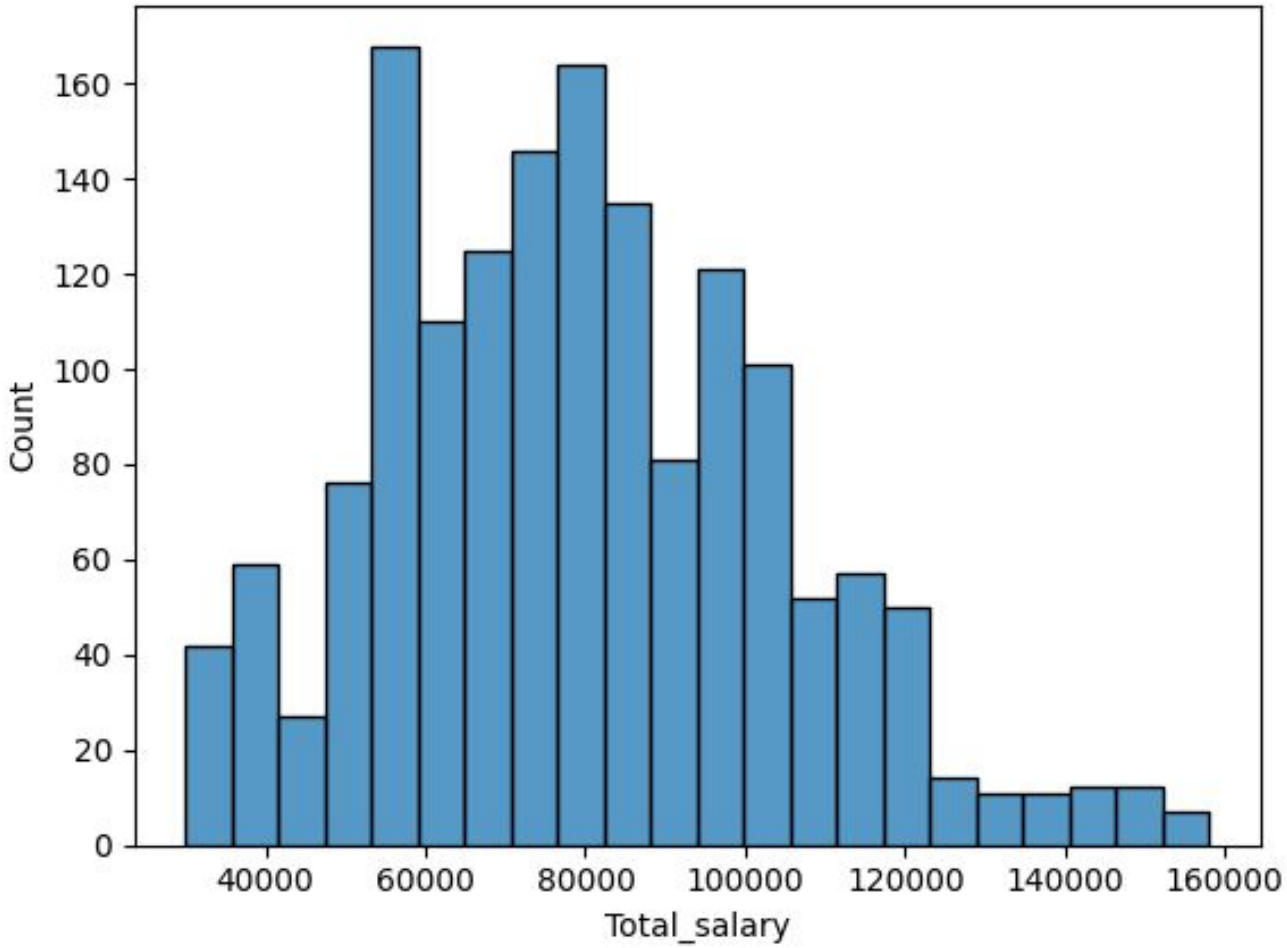
# EDA Results

## Univariate analysis: Partner’s salary (continued)



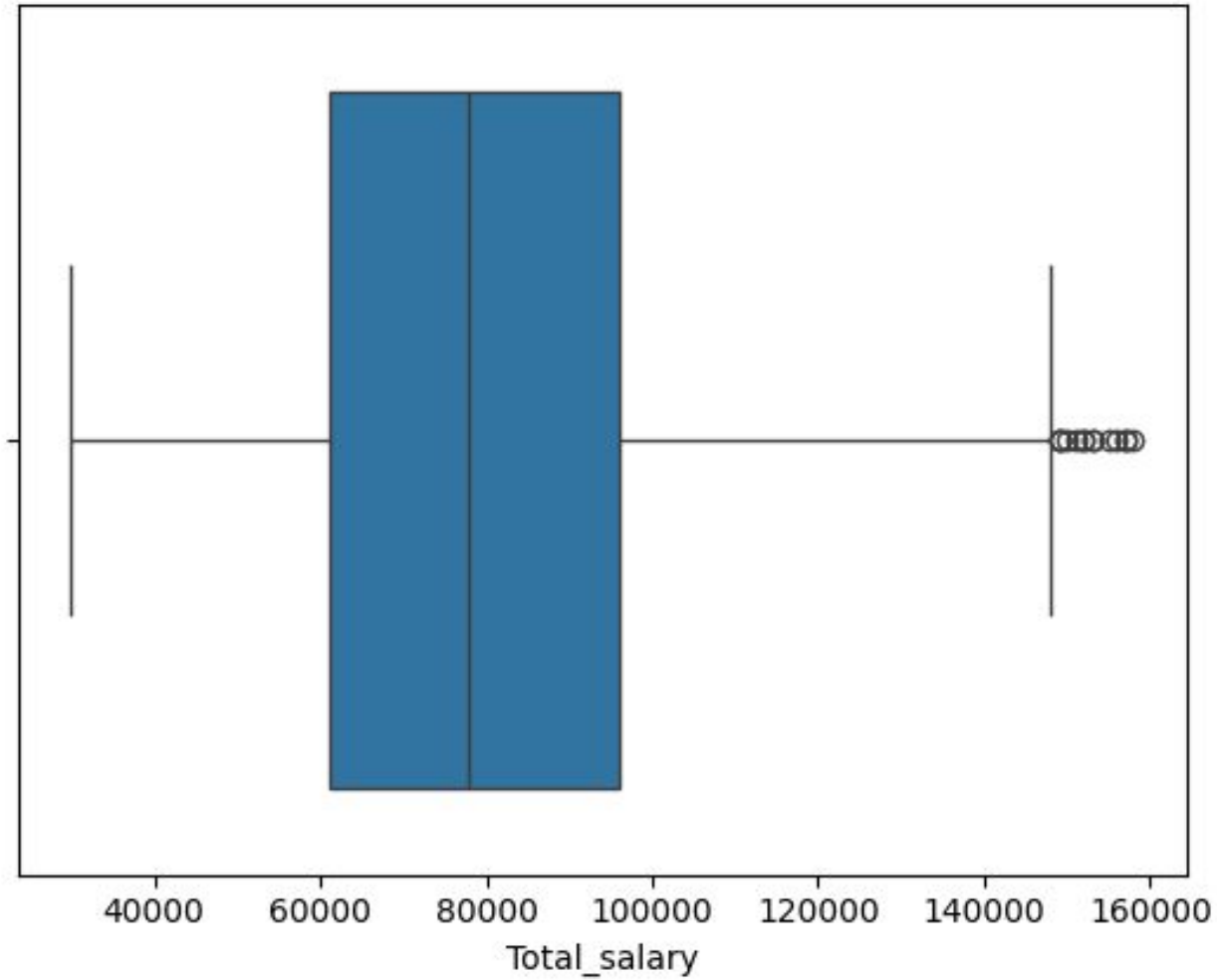

- There are no outliers
- The distribution is rightly skewed

# EDA Results
# Univariate analysis: Total salary




- The data does not follow a normal distribution
- Customers with total salaries of about 58000 and 80000 tend to buy cars more than others

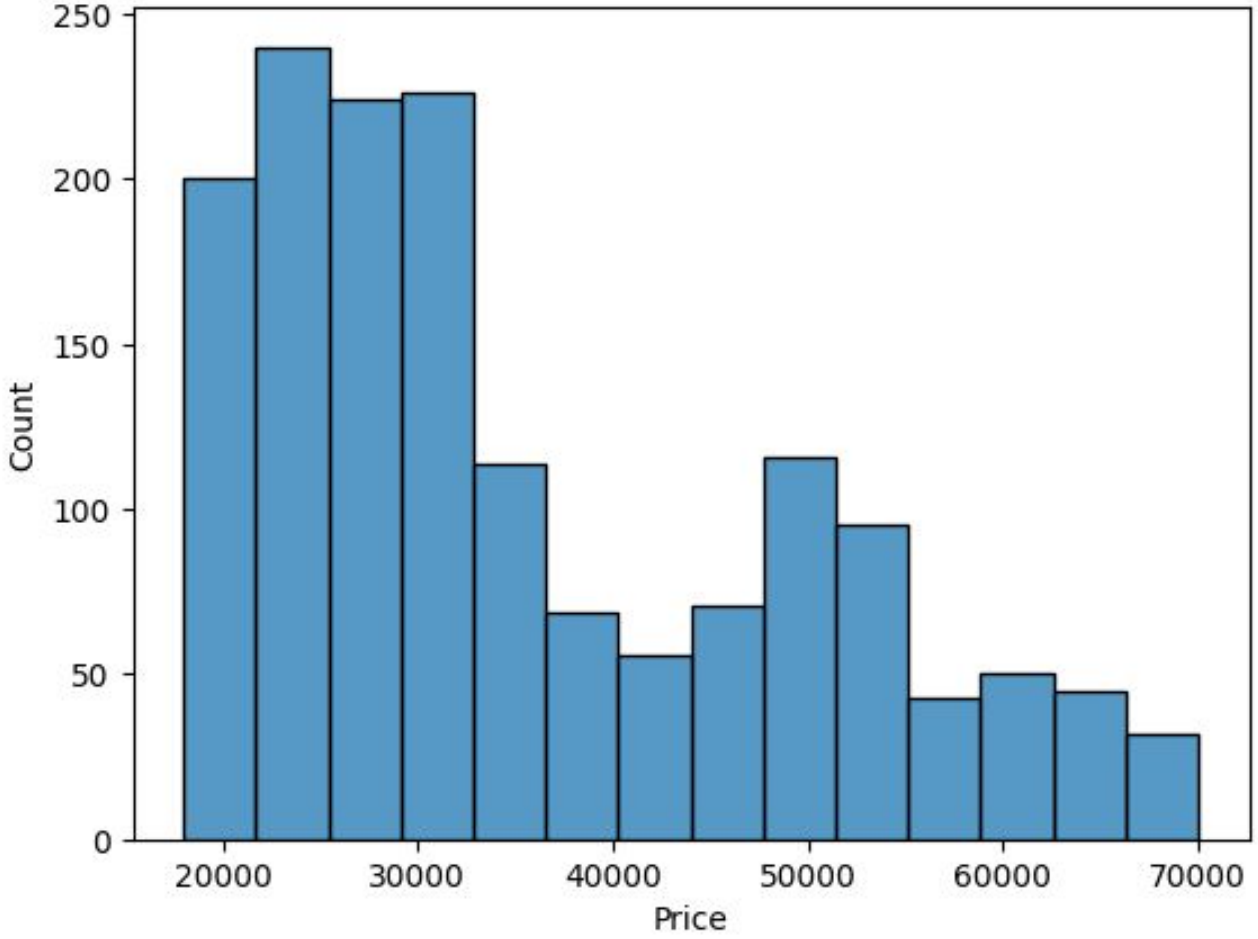
# EDA Results
## Univariate analysis: Total salary (continued)




- There are outliers
- The distribution slightly skewed to the right.
- The median total salary is around 79000

# EDA Results

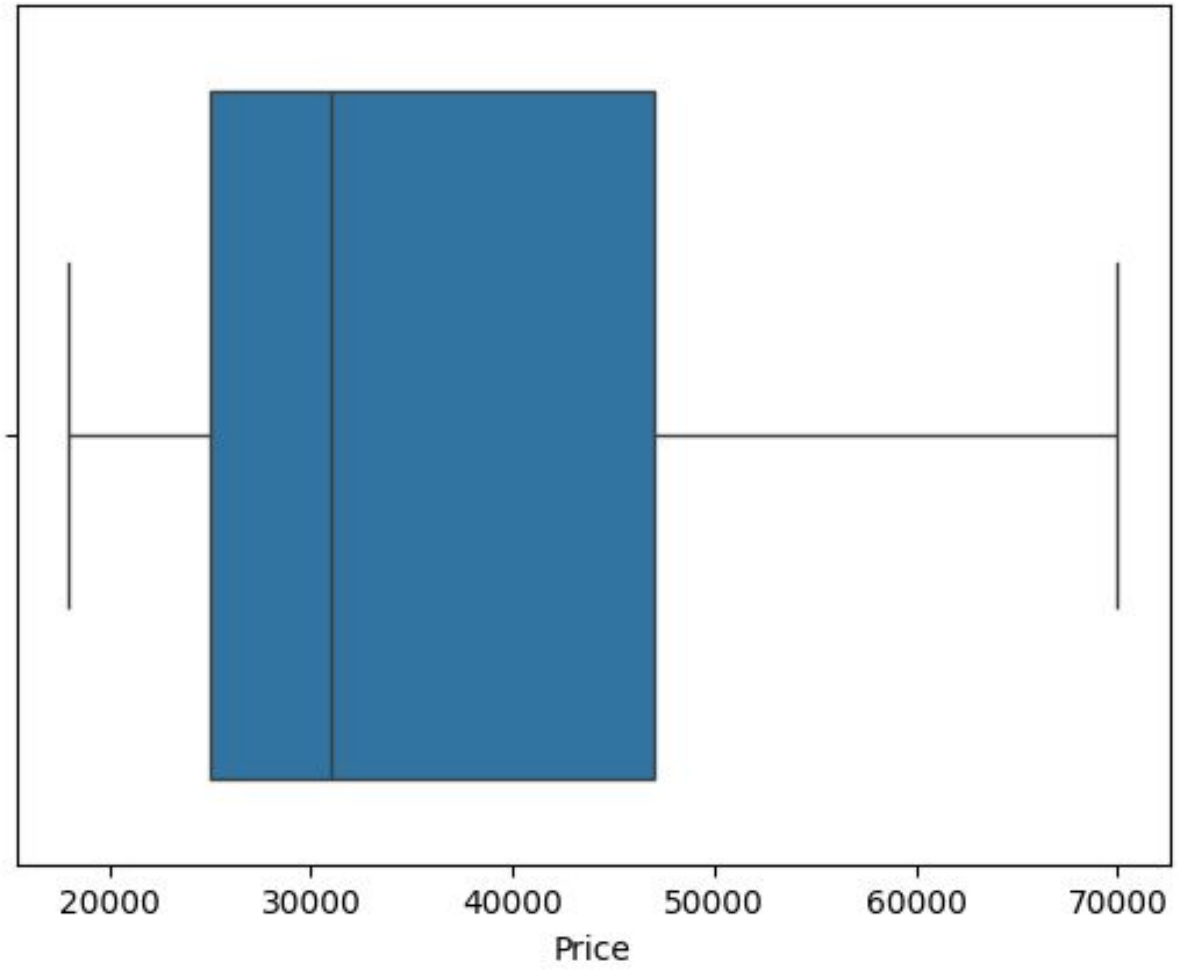
## Univariate analysis: Price




- Cars priced between 21000 and 25000 have more buyers
- Expensive cars of about 69000 to 70000 tend to attract least number of buyers

# EDA Results

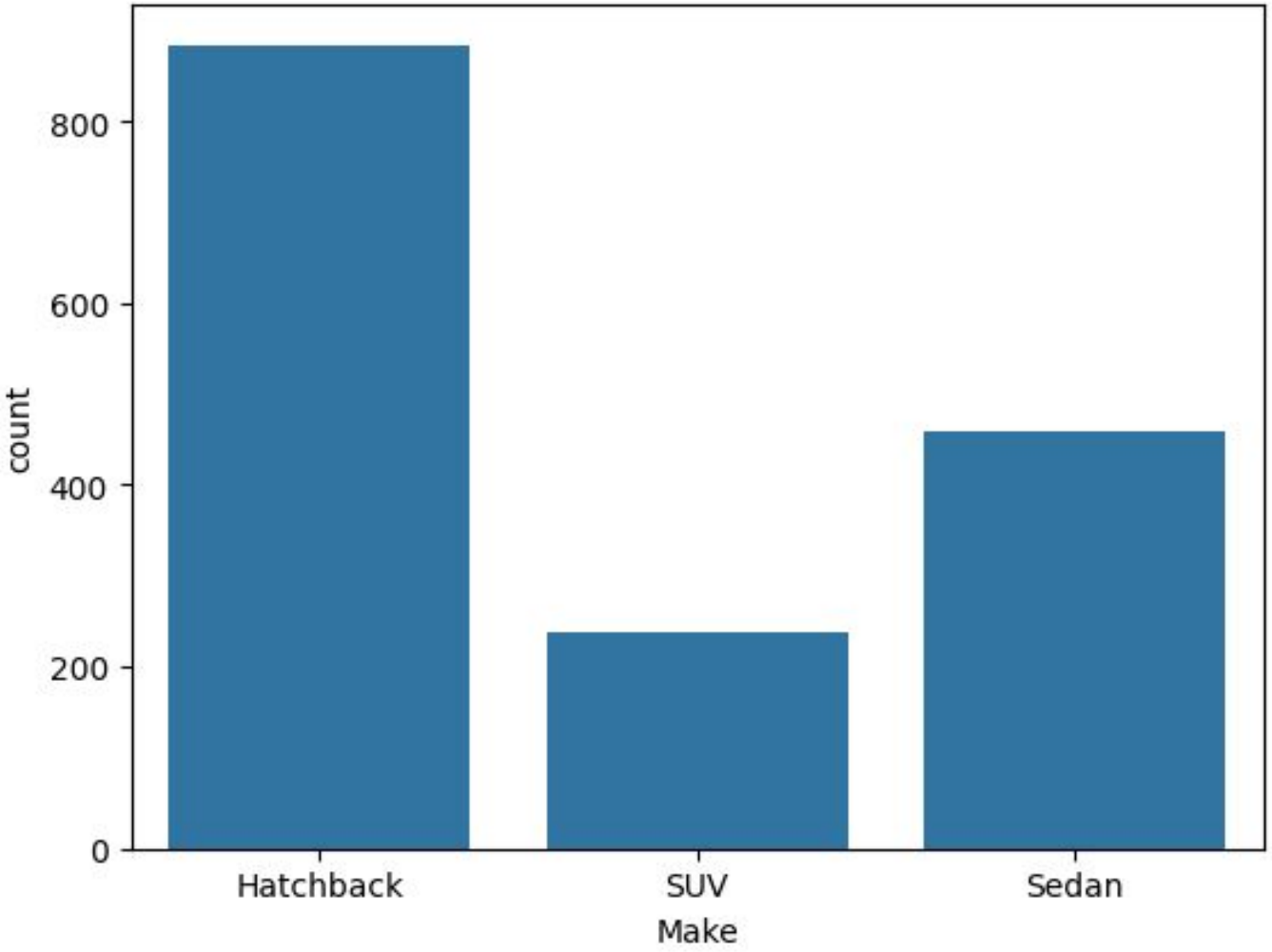
## Univariate analysis: Price (continued)




- There are no outliers
- The median price is around 31000
- The distribution is rightly skewed

# EDA Results

## Univariate analysis: Make



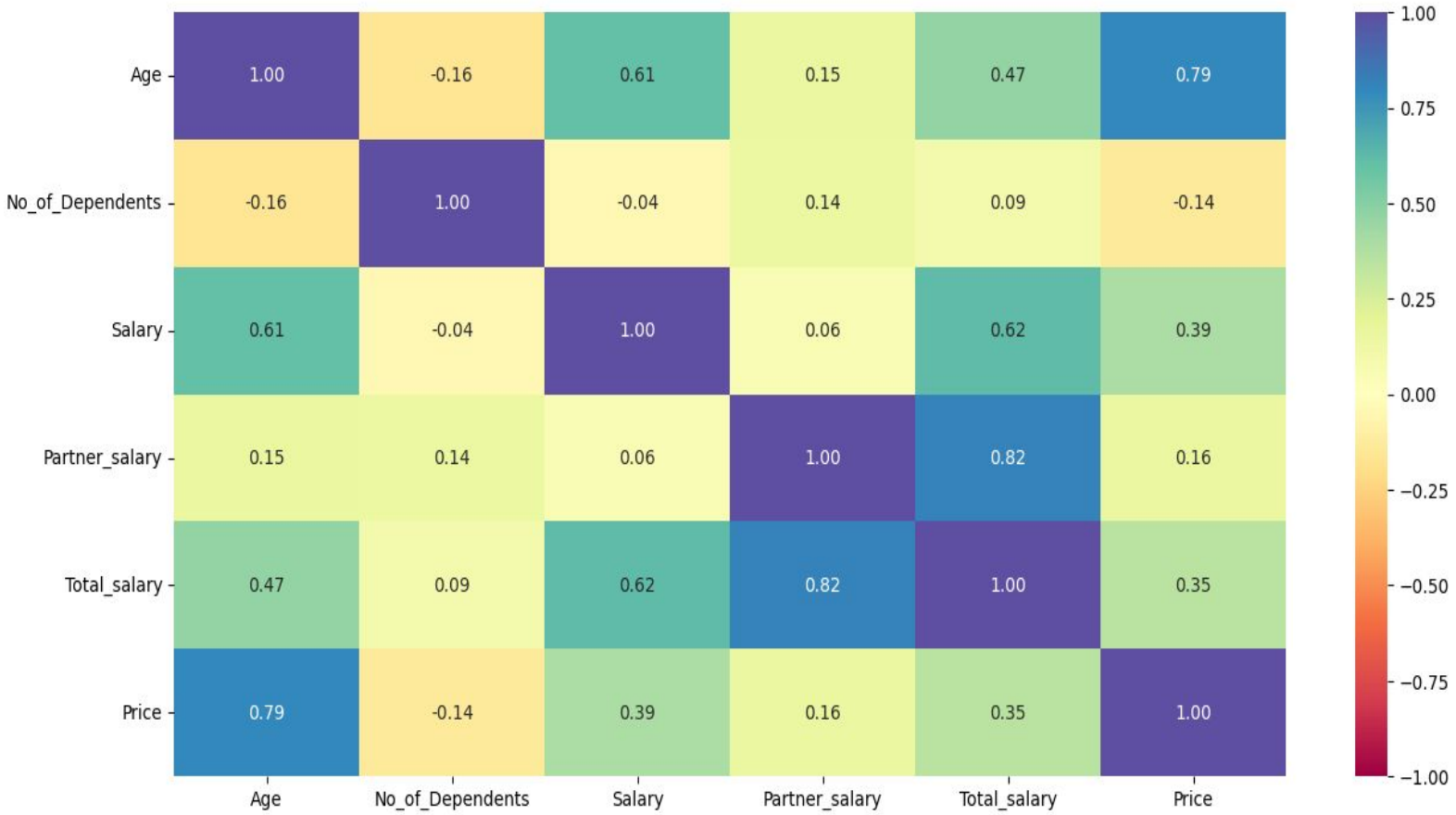

- There are three car types (coded with the variable “Make”) in the data: Hatchback, SUV and Sedan
- Hatchback has the highest count meaning that it was bought by most customers
- SUV has the least count, it has the least number of buyers
- About 400 customers bought Sedan

# EDA Results
## Univariate analysis

- There are 444 cars of make Hatchback and priced above 25000.
- Number of owners who have purchased cars pricing higher than their salary: 96
- Number of owners who have taken personal loan and have cars pricing higher than their salary: 44

# EDA Results

## Bivariate analysis: Heat map of numerical variables




The following variables are strongly positively correlated:

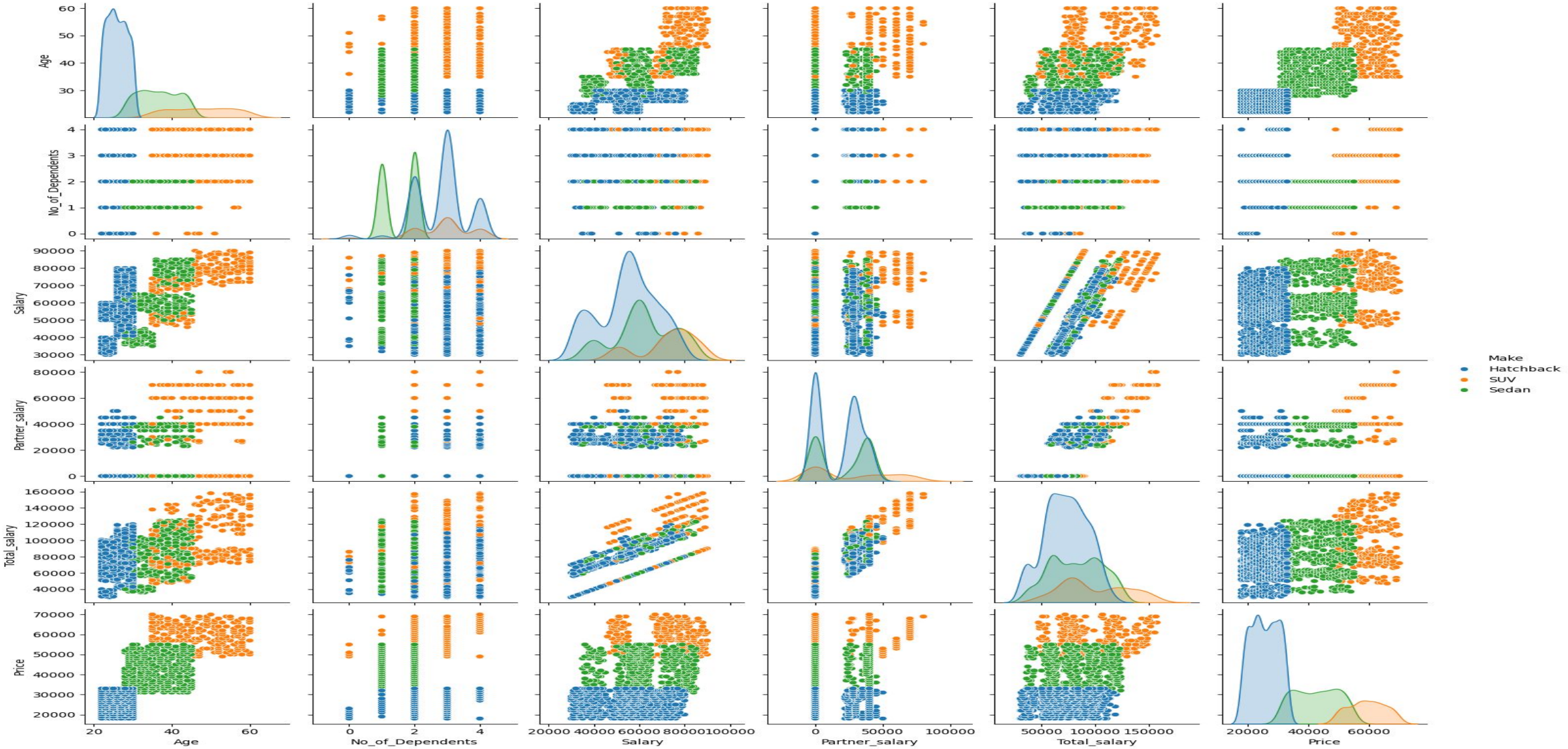
- Age and Price
- Partner salary and Total salary

There is a negative correlation between the following variables:

- Price and Number of dependents
- Age and Number of dependents
- Salary and Number of dependents

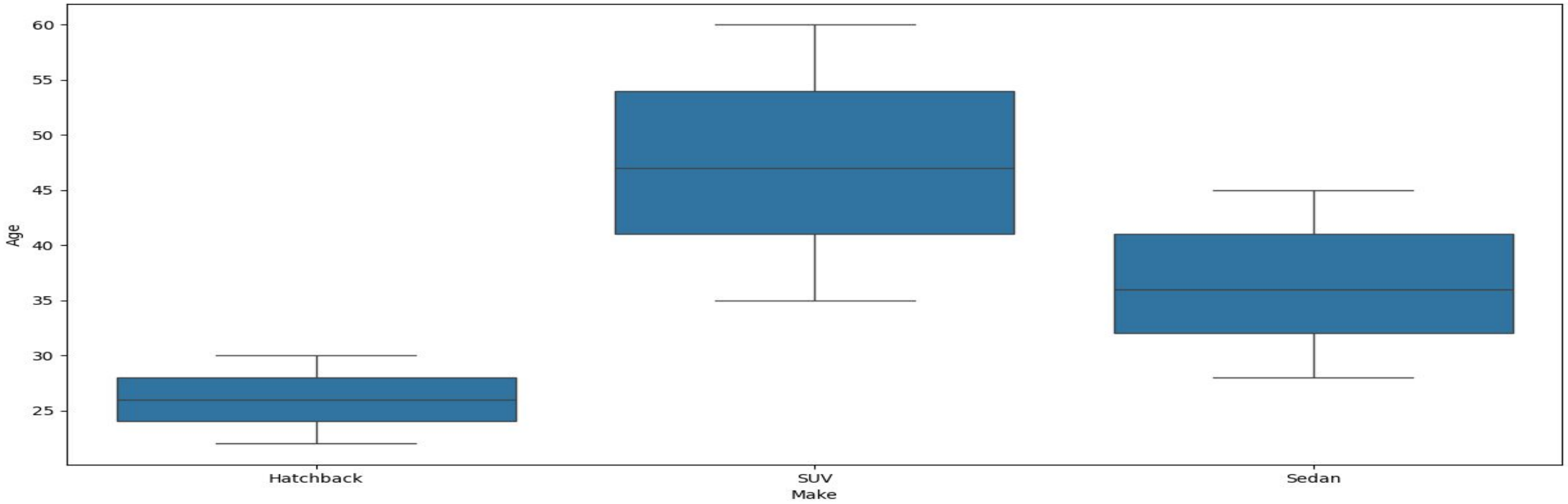
# EDA Results

## Multivariate analysis: Pairplot of numerical variables



# EDA Results

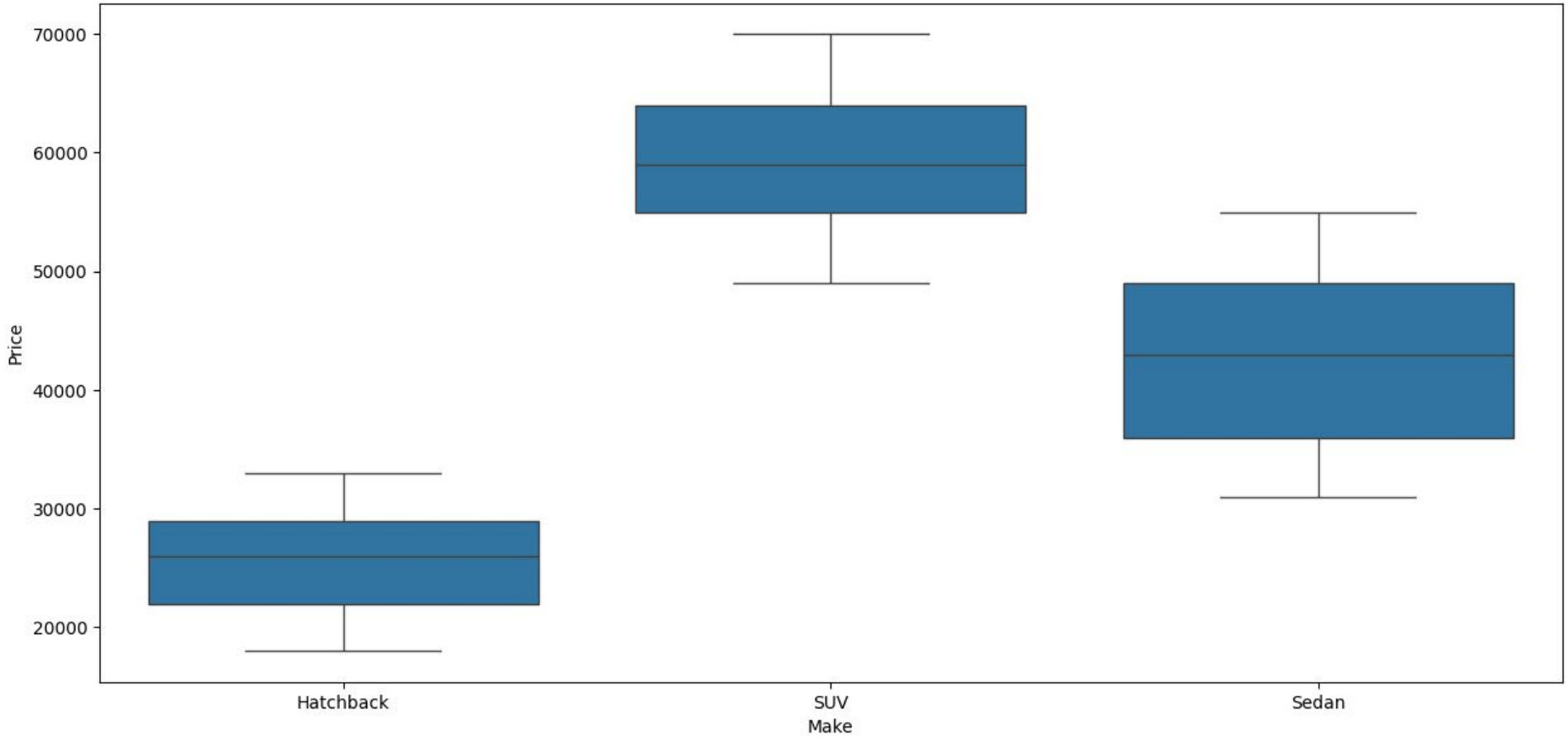
## Multivariate analysis: Boxplot of Age vs Make




- The boxplot shows that when cars are compared based on their Make, SUV was most purchased car by older (in terms of age) customers

# EDA Results

## Multivariate analysis: Boxplot of Price vs Make




Please see comments on next slide

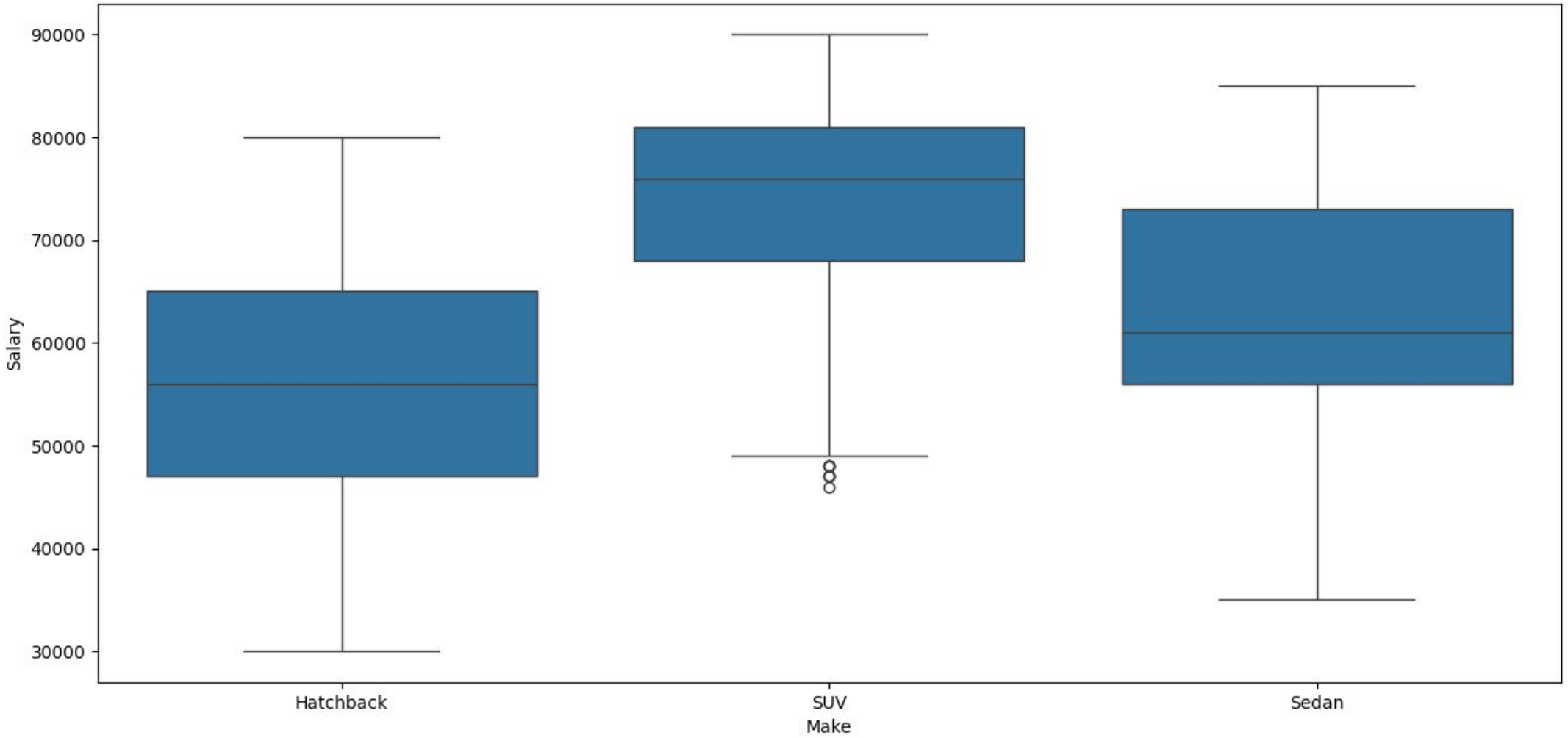
# EDA Results

Multivariate analysis: Boxplot of Price vs Make

- The boxplots above show that SUV cars are generally more expensive, and have a median price of about 60000
- SUV cars cost up to about 70000
- Hatchback cars are the least expensive then followed by Sedan cars

# EDA Results

Multivariate analysis: Boxplot of Salary vs Make




Please, see next slide for comments

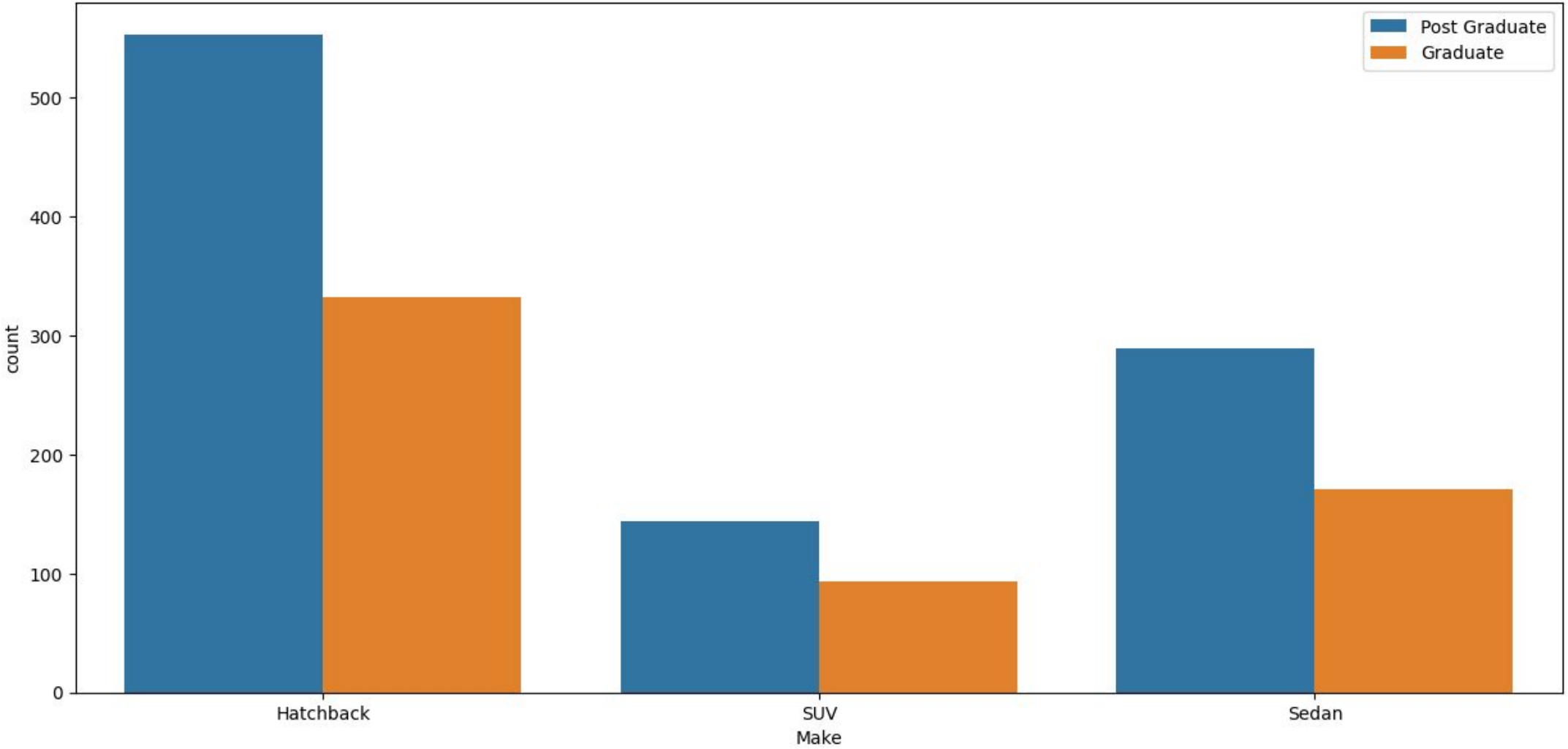
# EDA Results

## Multivariate analysis: Boxplot of Salary vs Make

- There are outliers meaning that there are some SUV cars purchased by customers with relatively low salaries
- People who earn between 82000 and 90000 did not purchase Hatchback, perhaps they found it too cheap and not worthy of their financial status
- The median salary of customers who purchased Sedan is 60000
- Comparing the three car Make, customers with salaries as low as 30000 could afford Hatchback while customers with salaries around 90000 could afford to buy SUV

# EDA Results

Multivariate analysis: Countplot of Make vs Education




Please see next slide for comments

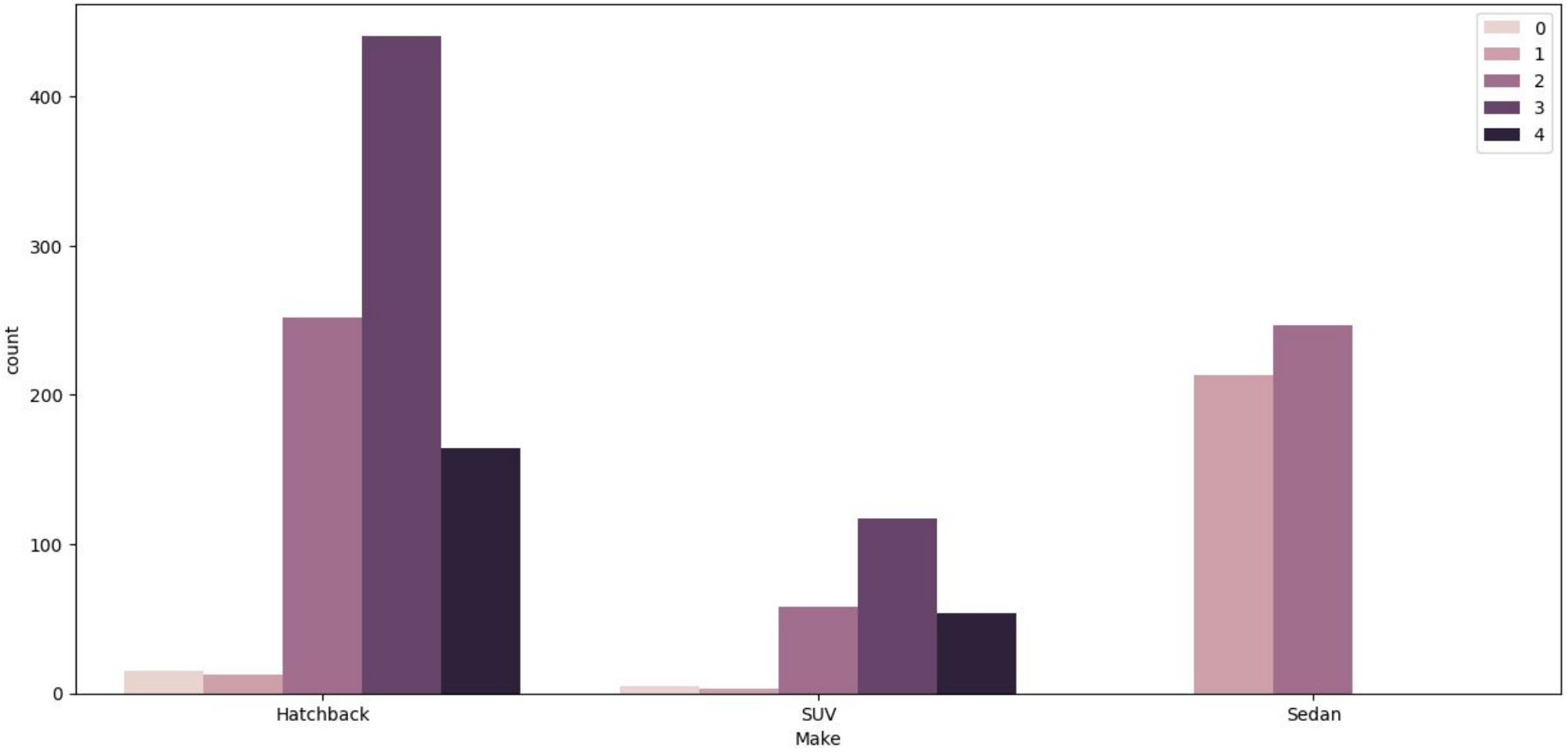
# EDA Results

Multivariate analysis: Countplot of Make vs Education

- Generally speaking, people with postgraduate degrees tend to be the ones who buy more cars. It will be good to know if people's level of education correlates with their salary level and if this in turn affects their propensity to buy cars (people's taste when it comes to purchasing cars).
- Hatchback car attracted the most postgraduate customers, while SUV attracted the least postgraduate customers
- Also, Hatchback car attracted the most undergraduate customers, while SUV attracted the least undergraduate customers

# EDA Results

## Multivariate analysis: Countplot of Make vs Number of Dependents




- Please see next slide for comments

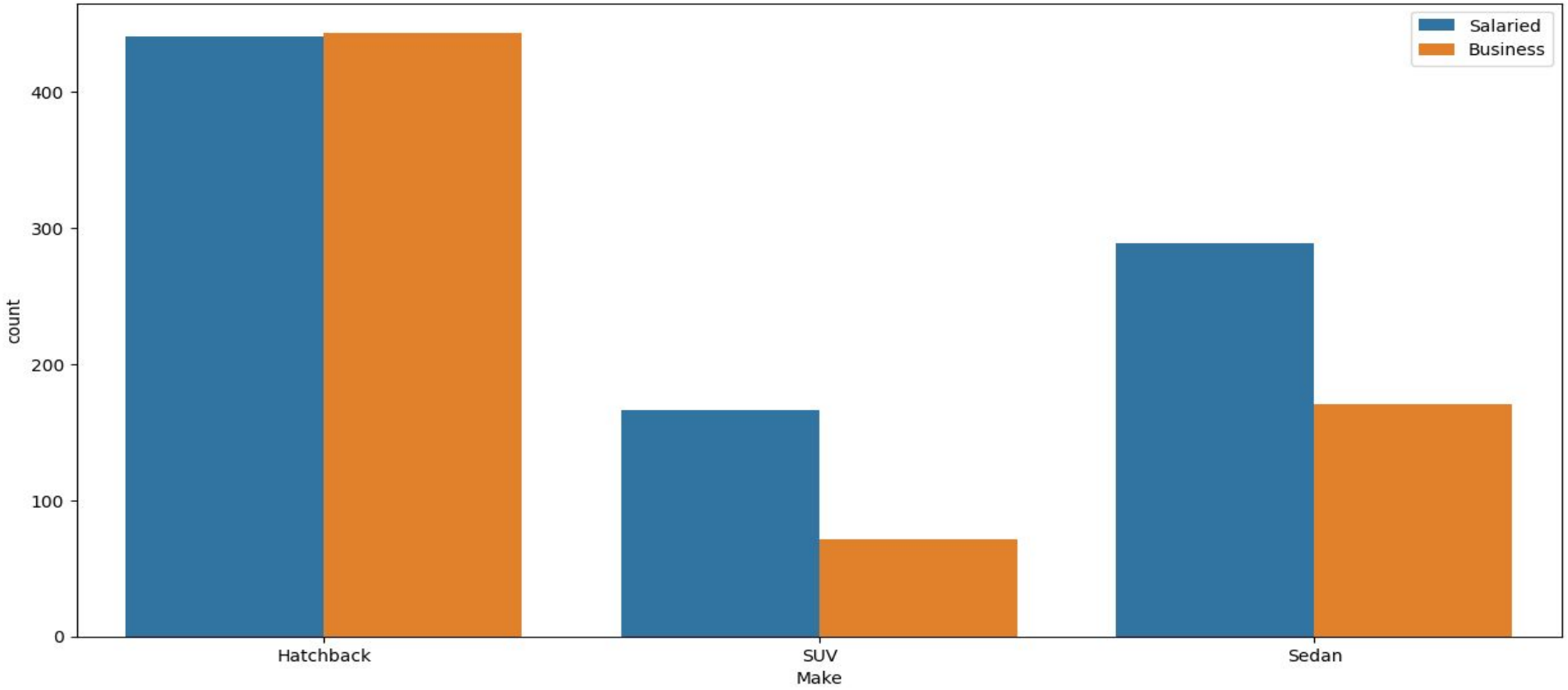
# EDA Results

Multivariate analysis: Countplot of Make vs Number of Dependents

- Generally, Customers with dependents bought Hatchback more than SUV
- Customers who buy Hatchback mostly have 3 dependents
- There are about 160 customers with 4 dependents who buy Hatchback
- The distribution of SUV buyers across No_of_dependents is skewed to the left
- Most of the people who bought SUV have 3 dependents
- People who bought Sedan either have 1 or 2 dependents

# EDA Results

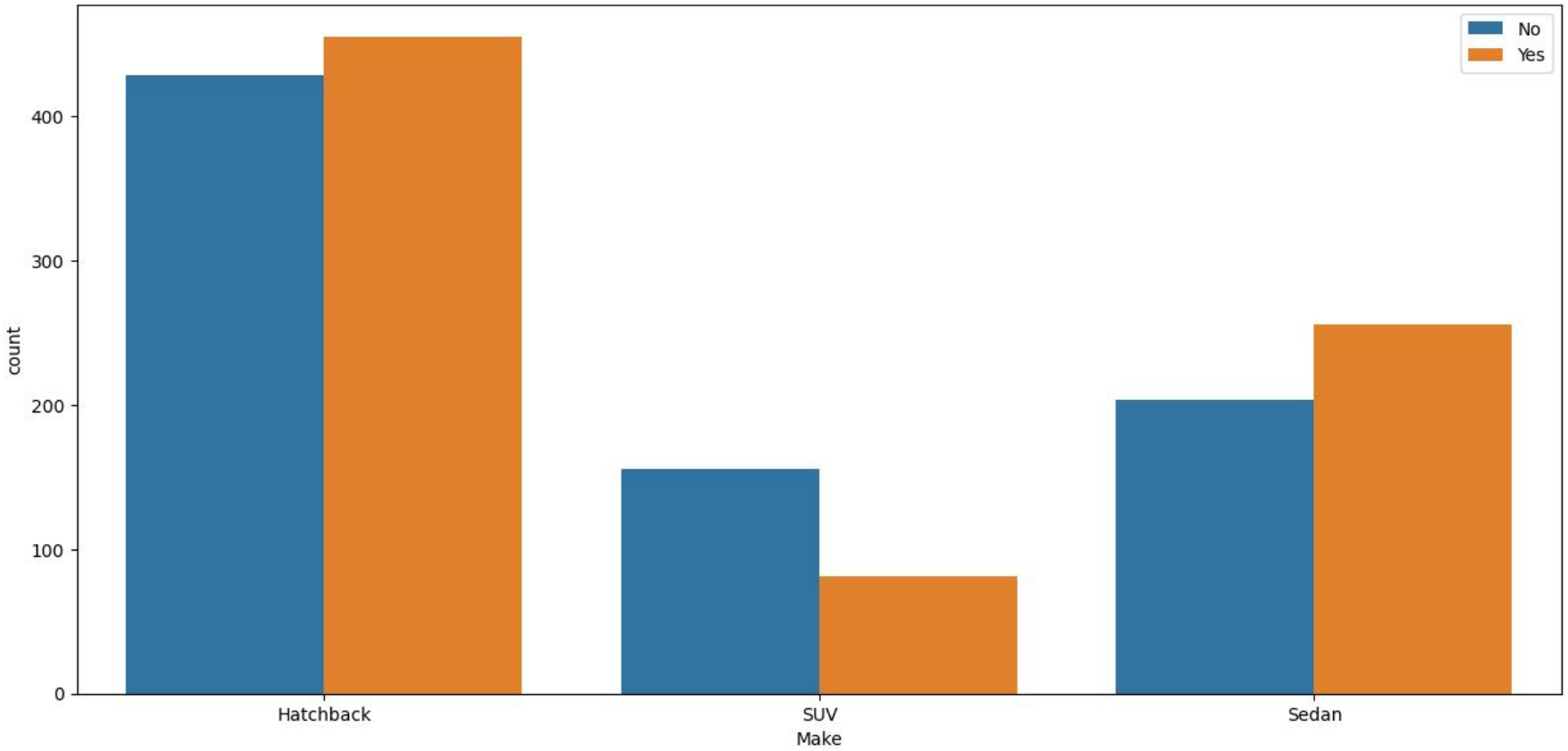
## Multivariate analysis: Countplot of Make vs Profession




- When it comes to whether customers earn salaries or do business, Hatchback seems to have approximately equal counts for both groups. That is, the probability that a customer who buys Hatchback is also a business person is 0.5.
- There are more salaried customers than business people who buy SUV
- There are more salaried customers that buy Sedan than business people

# EDA Results

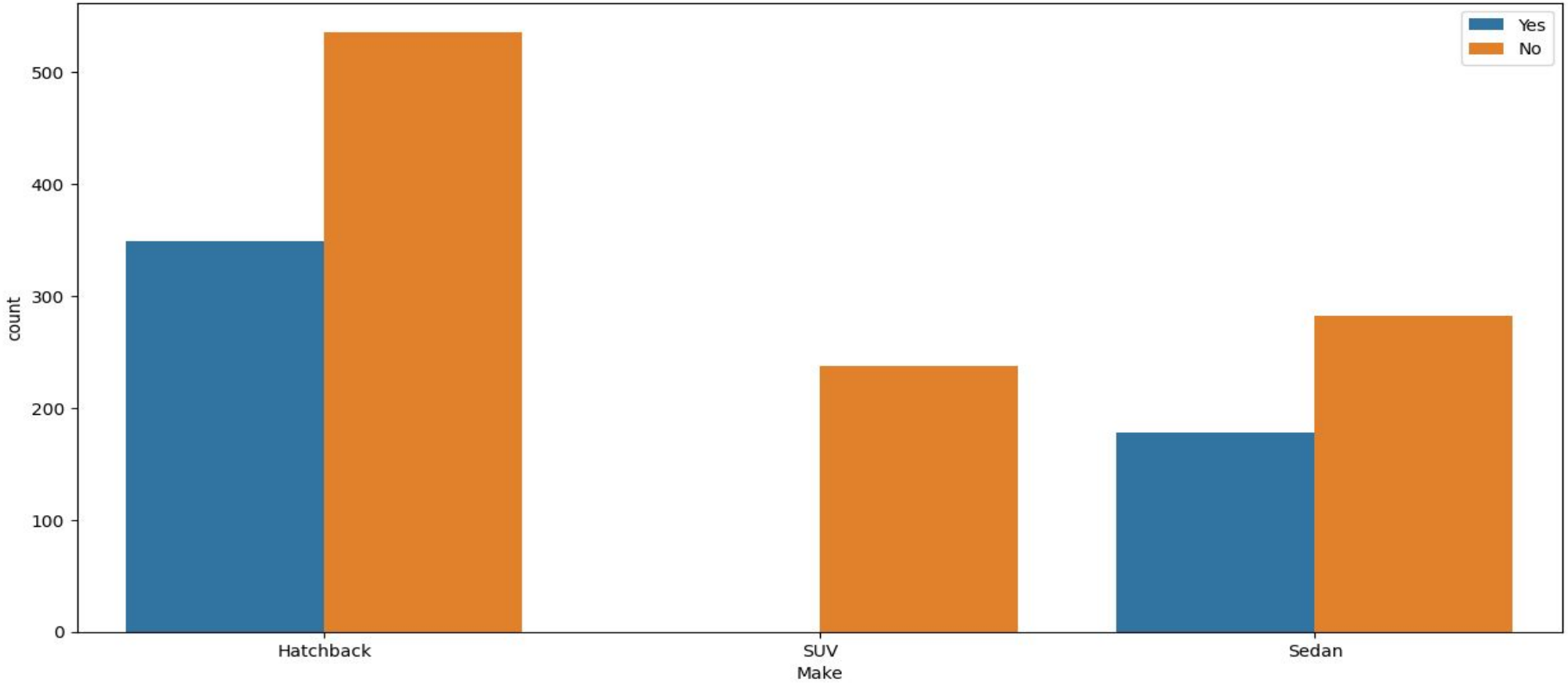
## Multivariate analysis: Countplot of Make vs Personal loan




- There are more customers who availed personal loan to buy Hatchback than those who did not.
- More customers bought SUV without personal loan. SUV cars are expensive, and it is interesting to see that more people bought it without personal loan.
- More people bought Sedan with personal loan

# EDA Results

Multivariate analysis: Countplot of Make vs House loan




- There are more people who buy Hatchback who do not have house loan than people who have house loan
- Only people without having house loan buy SUV. It seems people who are not paying loan have the financial capacity to own SUV.
- There are more people without house loan who buy Sedan
- Generally, people who have house loan tend to not buy cars

# EDA Results

Grouping data w.r.t to car types to build customer profiles

Grouping data w.r.t to Hatchback

- The mean age of customers who buy Hatchback is 25.874434 while the median age is 26
- Most customers who buy Hatchback are male, are business people, are married, have post graduate degrees, have personal loan, but not house loan, and have partners who work
- The mean salary of customers who buy Hatchback is 54969.46 while the median salary is 56000, the minimum salary is 30000. This shows that Hatchback is pretty much affordable for customers.

# EDA Results

Grouping data w.r.t to car types to build customer profiles

Grouping data w.r.t to Sedan

- The mean age of customers who buy Sedan is 36.46 while the median age is 36
- Most customers who buy Sedan are male, are salaried people, are married, have post graduate degrees, have personal loan, but not house loan, and have partners who work
- The mean salary of customers who buy Sedan is 62213.04 while the median salary is 61000, the minimum salary is 35000. This shows that Sedan is pretty much affordable for customers.

# EDA Results

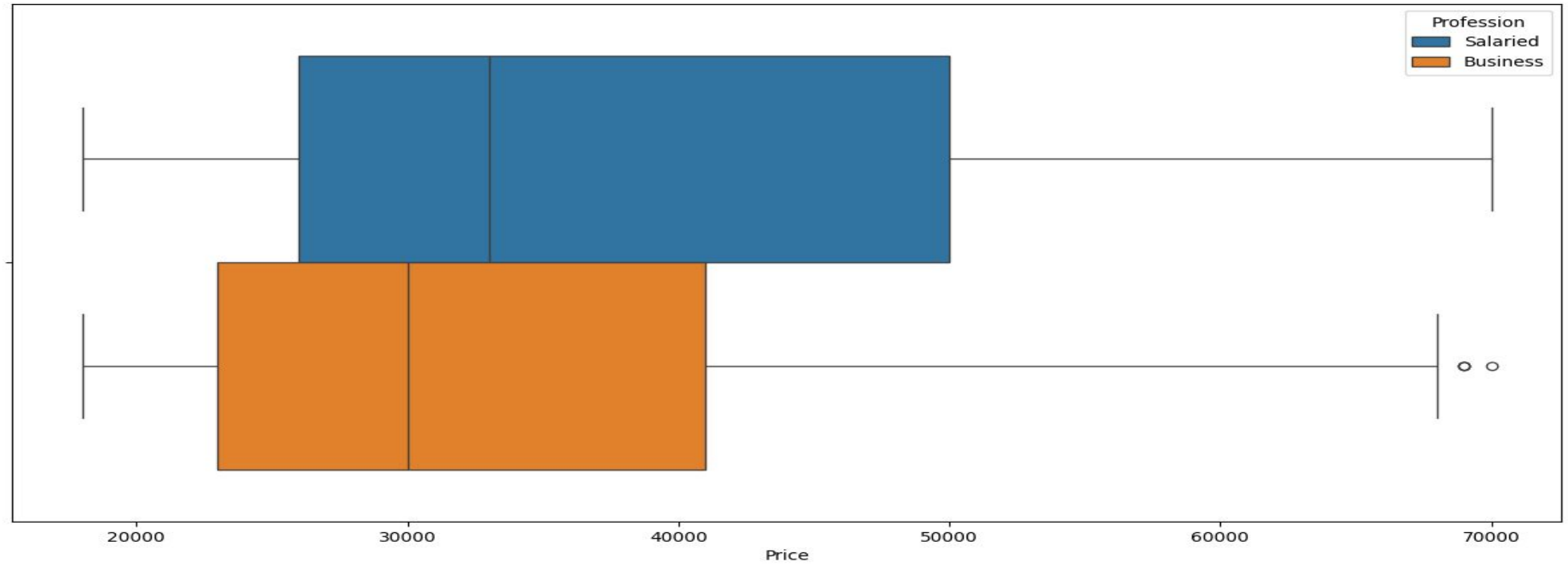
Grouping data w.r.t to car types to build customer profiles

Grouping data w.r.t to SUV

- The mean age of customers who buy SUV is 47.6 while the median age is 47
- Most customers who buy SUV are male, are salaried people, are married, have post graduate degrees, have no personal loan, no house loan, and have partners who work.
- The mean salary of customers who buy SUV is 72683.524 while the median salary is 76000, the minimum salary is 46000. This shows that SUV seems to be more expensive.

# Customer Segmentation

## Profession vs Price

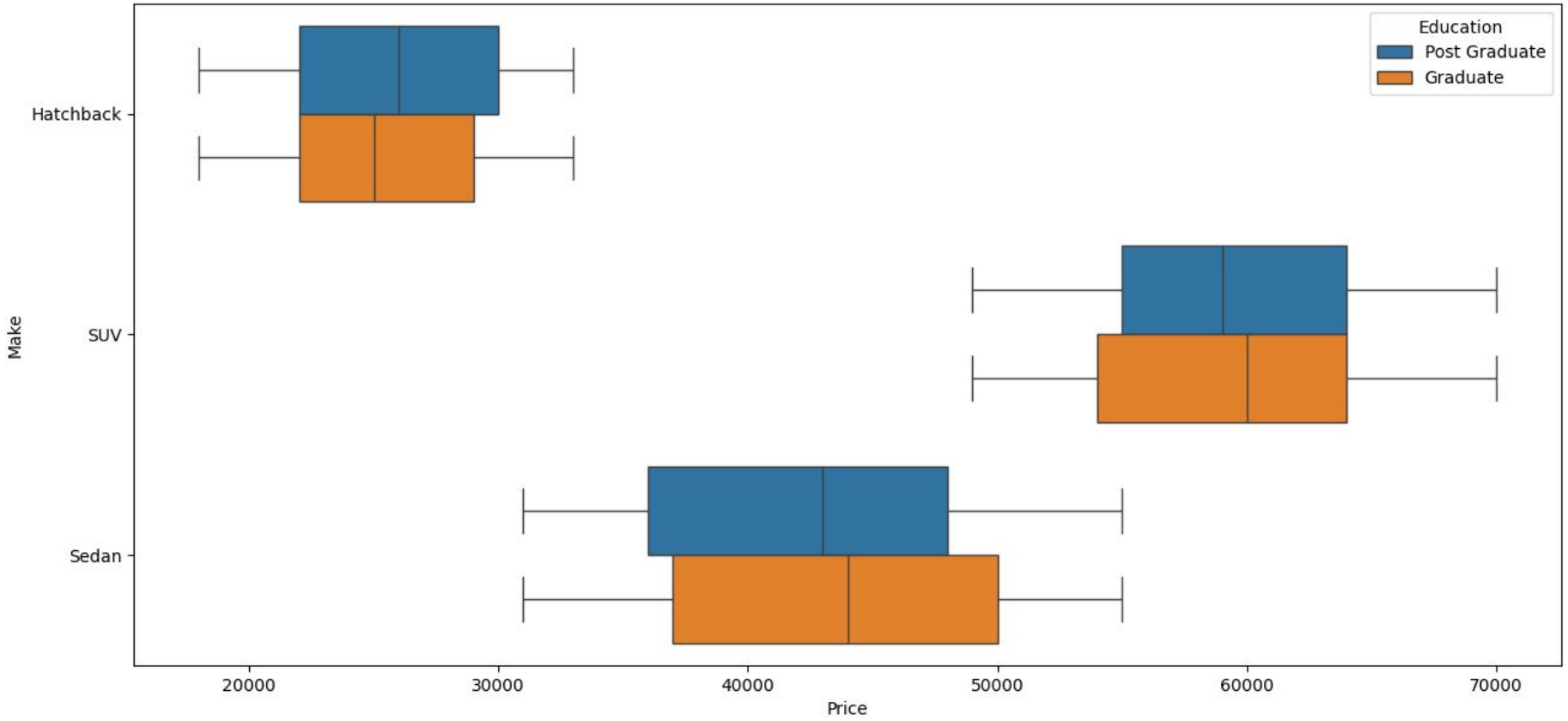



Although people who do business generally tend to buy cheaper cars, but there are two customers who bought two expensive cars close to 70000 in price

- Generally, salaried people tend to buy more expensive cars
- 75 percent of business people bought cars priced between 19000 and 41000
- 50 percent of salaried people bought cars priced between 32000 to 70000

# Customer Segmentation

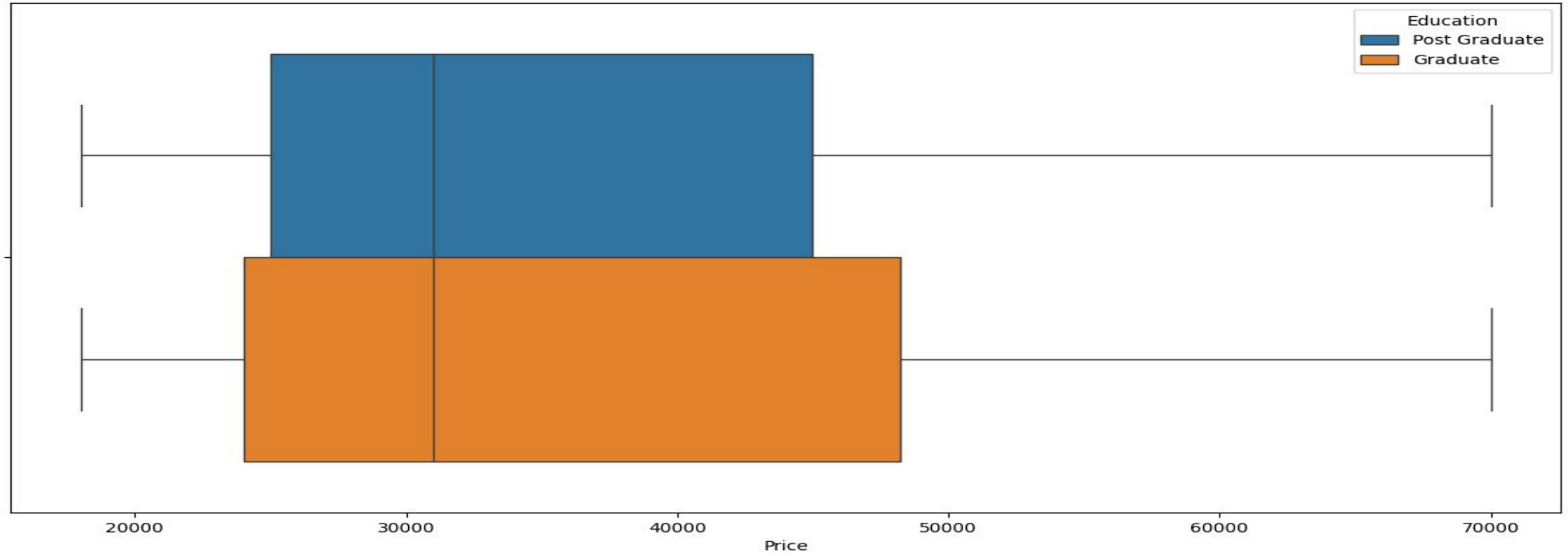
## Education vs Price vs Make




- The median price of Hatchback cars bought by customers with postgraduate degrees is more than the median price of cars bought by those with undergraduate degrees
- The reverse is the case for Sedan cars, the median price of Sedan cars bought by customers with postgraduate degree is higher.

# Customer Segmentation

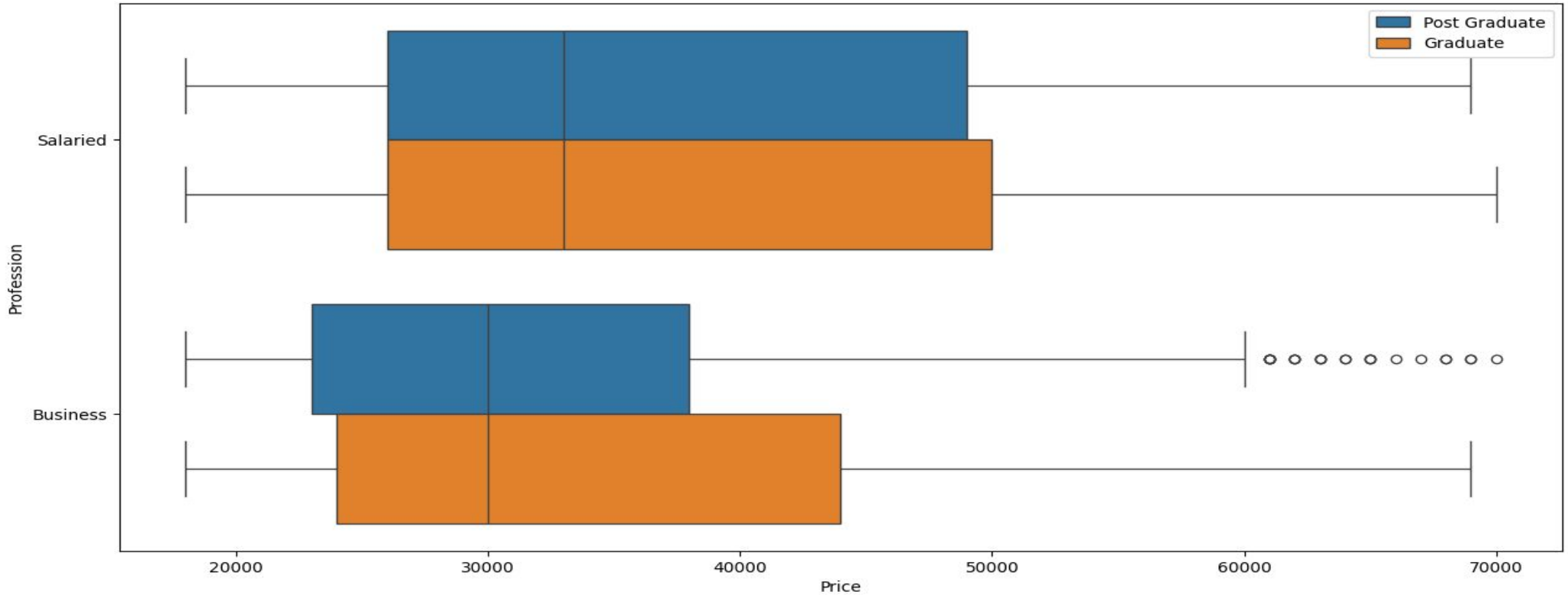
## Education vs Price




- The median price of cars bought by people with postgraduate degree holders is same as the median of cars bought by those with graduate degree

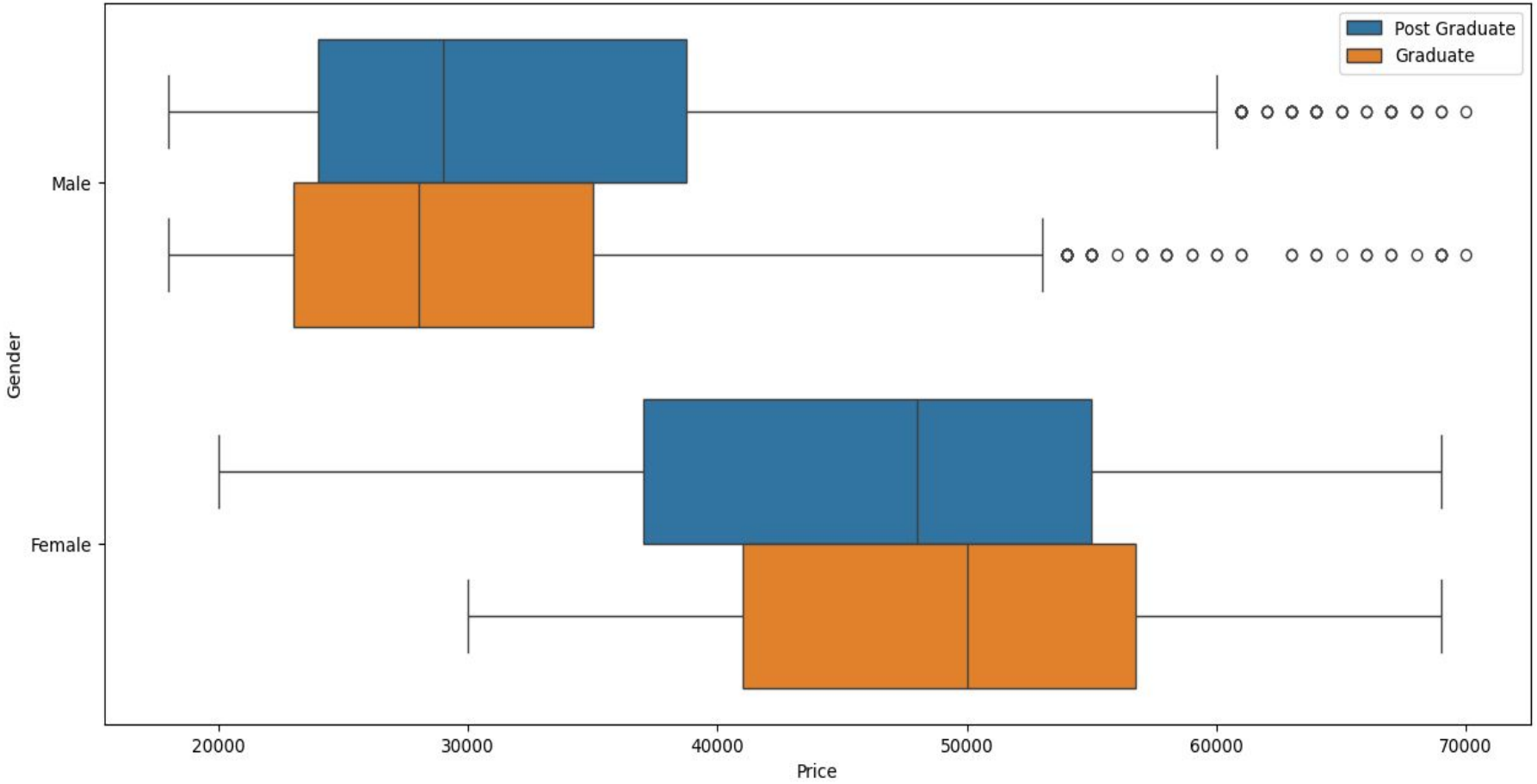
# Customer Segmentation

## Education vs Profession



# Customer Segmentation

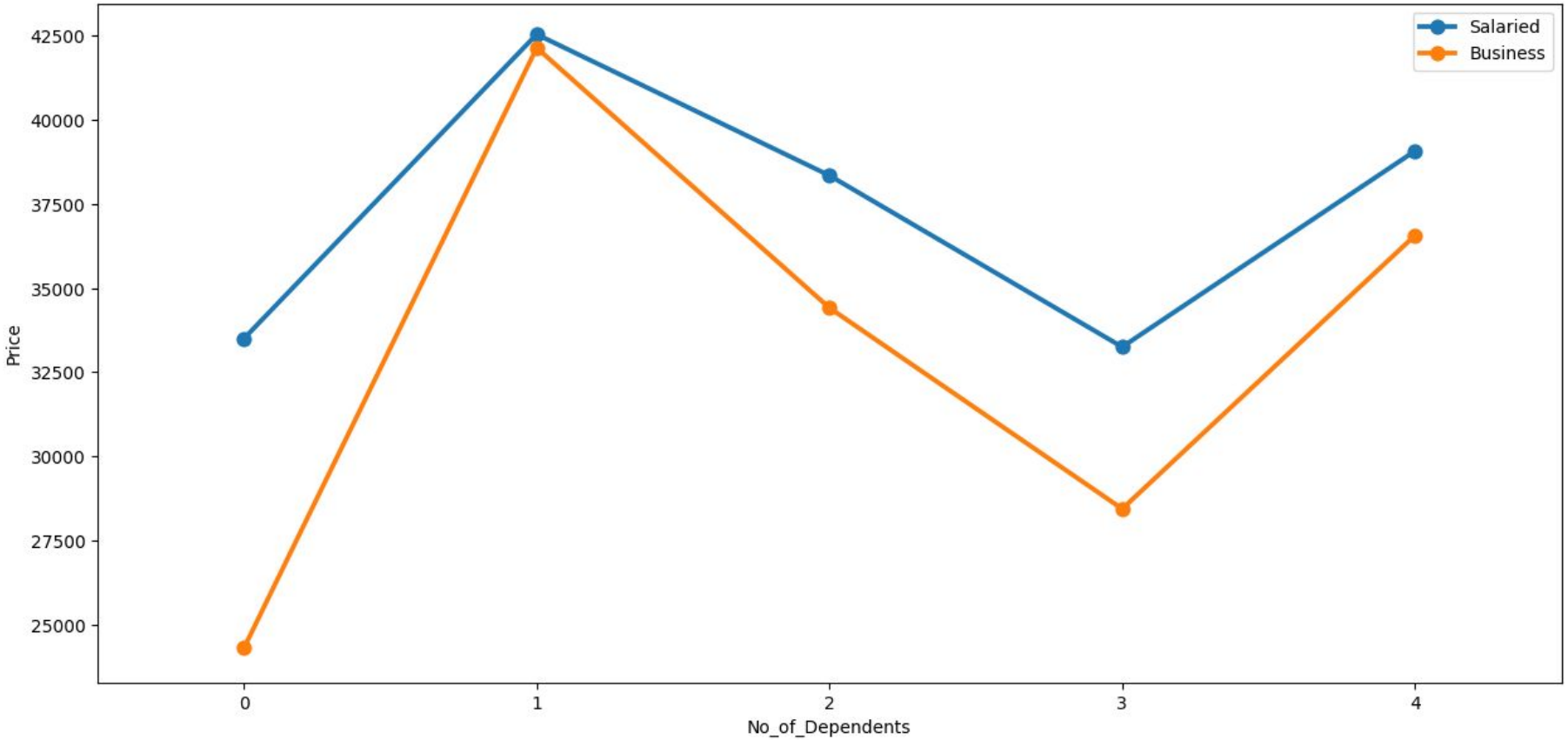
## Gender vs Price vs Education




- The first two boxplots shows there are outliers in the data

# Customer Segmentation

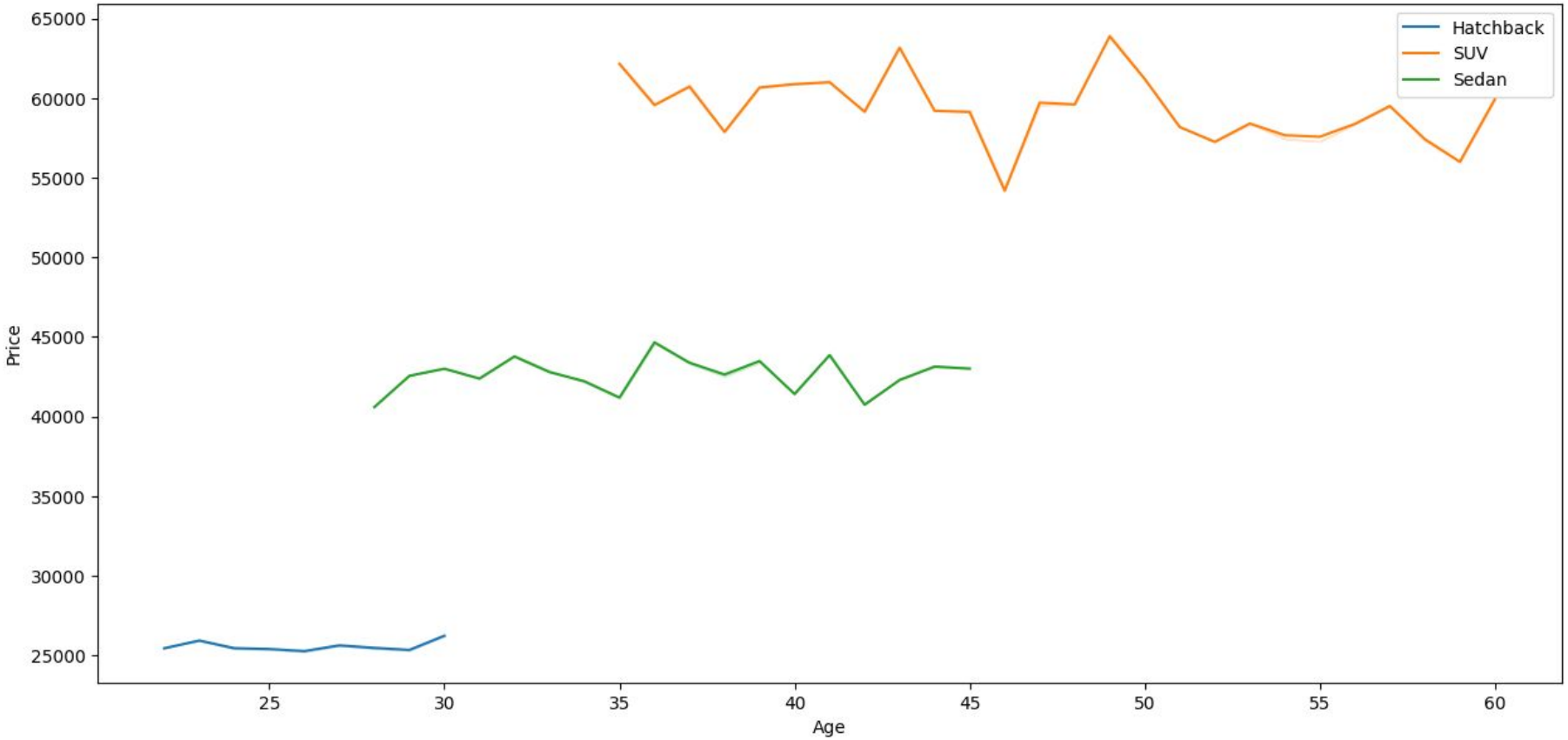
## Number of dependents vs Price vs Profession




- For any given number of dependents, salaried customers tend to buy higher priced cars than their business oriented counterparts.

# Customer Segmentation

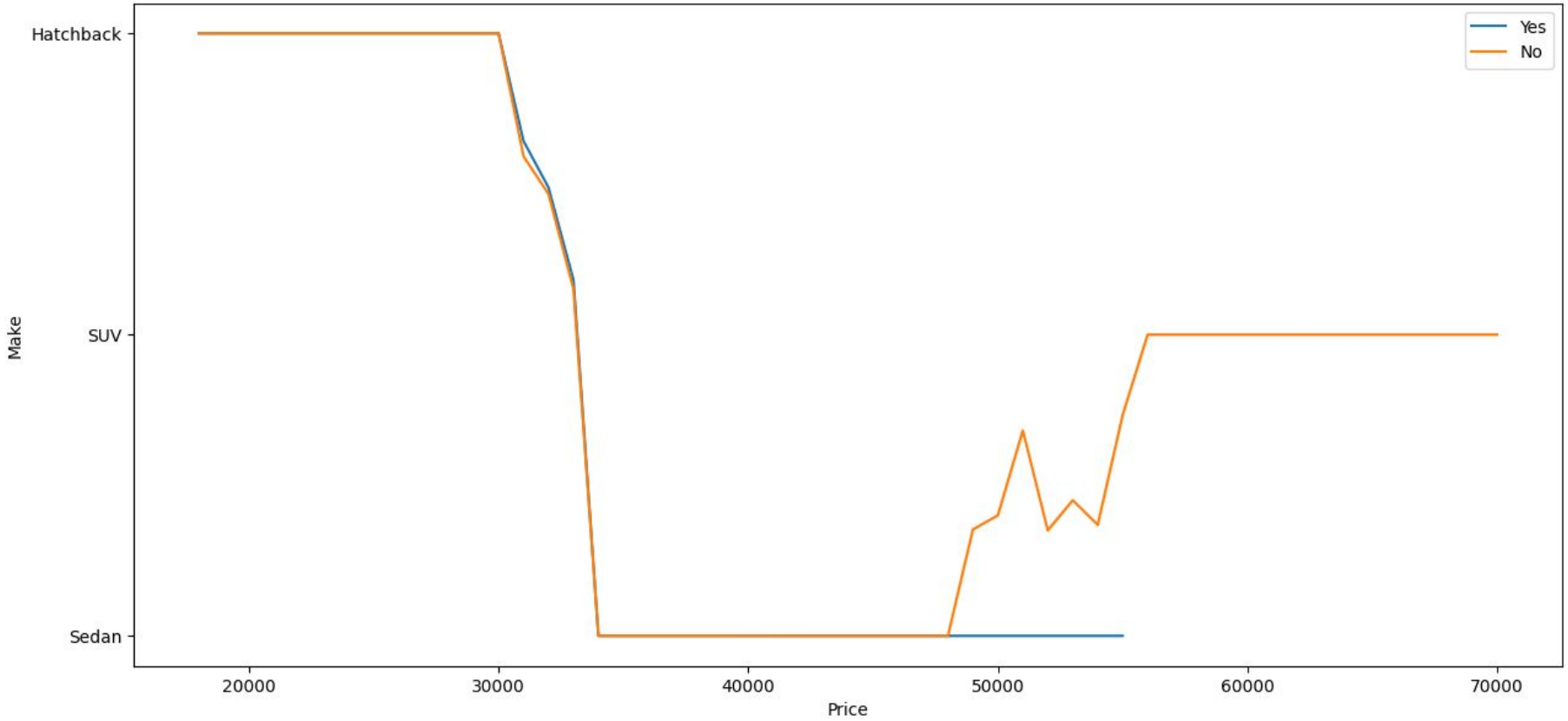
## Age vs Price vs Make




- Older customers tend to buy more expensive cars like SUV.

# Customer Segmentation

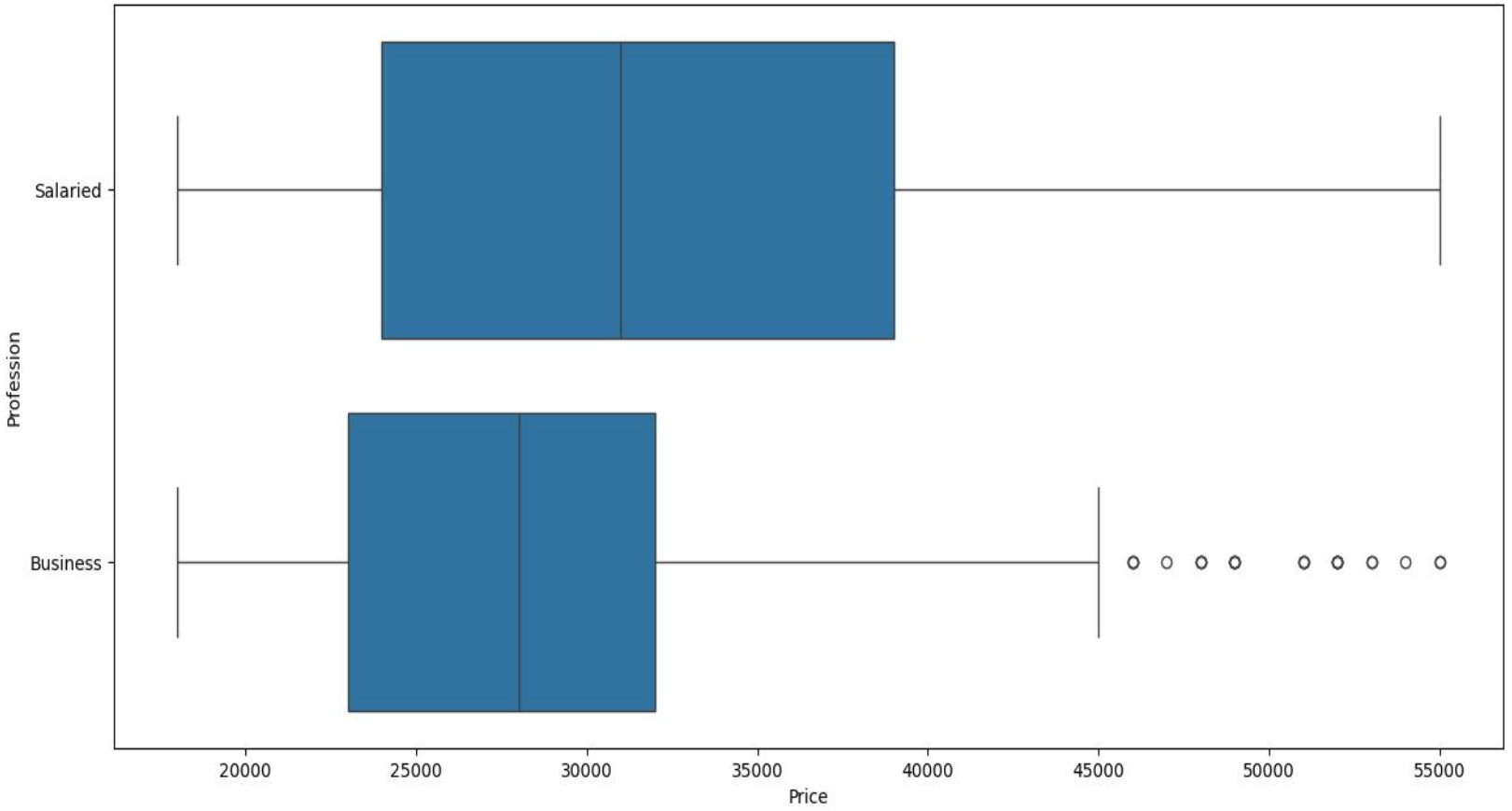
## Make vs Price vs House loan



# Average car price by profession

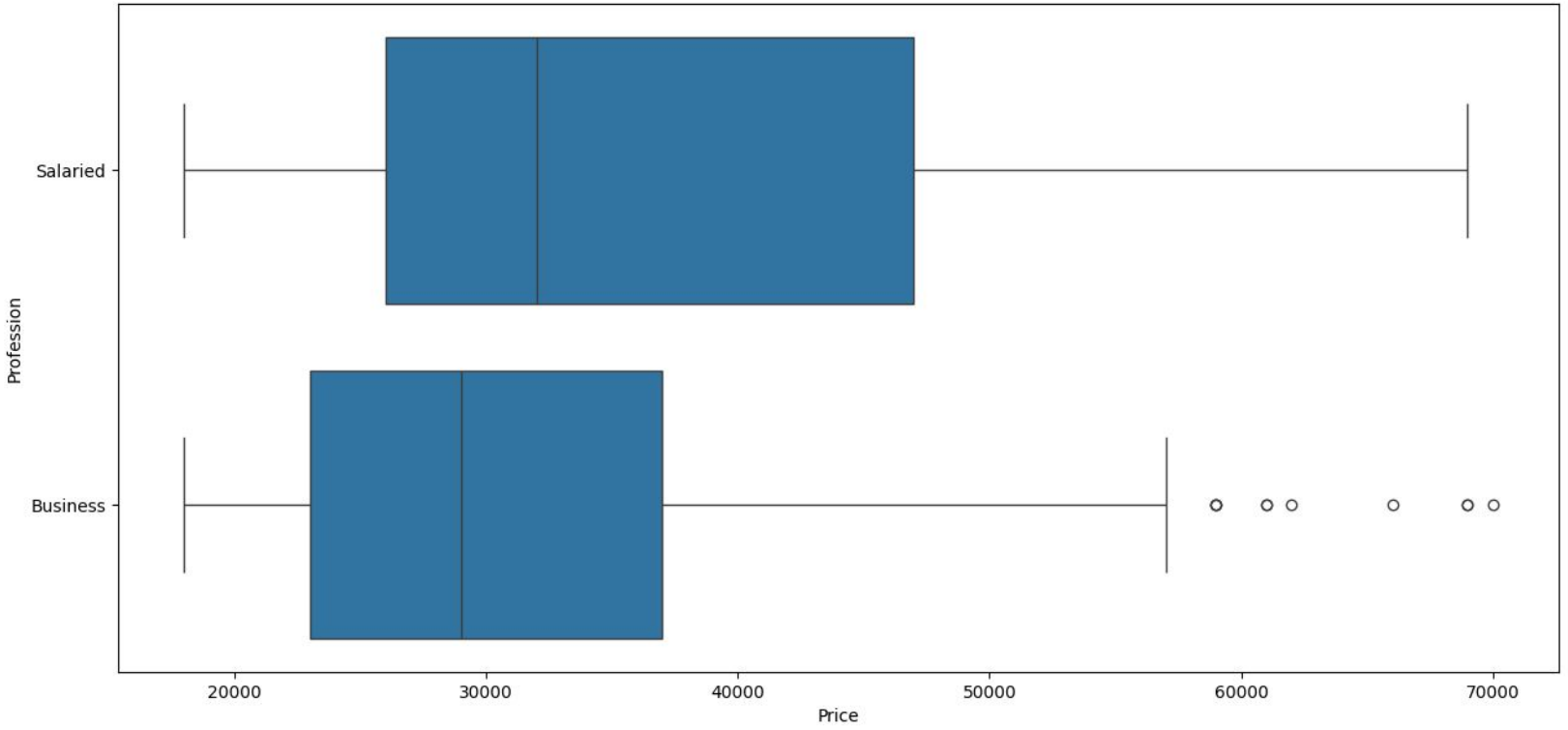
- Average car price purchased by salaried customers is 37036.18
- Average car price purchased by business oriented customer is 32830.22

# Price by Profession of customers with Home Loan




- There are some customers with home loans who are business oriented and unusually buy expensive cars around 47000 to 70000
- 50 percent of customers with personal loans buy cars priced from 32000 and above

# Price by Profession of customers with Personal Loan




- There are some customers with personal loans who are business oriented and unusually buy expensive cars around 60000 to 70000
- 50 percent of customers with personal loans buy cars priced from 32000 and above

# Conclusion

- Older people tend to prefer SUV the most, followed by Sedan and then Hatchback
- Most salaried customers prefer Hatchback, followed by Sedan and then SUV
- Females generally tend to spend more money on cars, there are a few men who buy expensive cars
- There is not much difference between postgraduates and undergraduate degree holders when it comes to the price of cars they buy.
- Hatchback cars are more preferred by customers with dependents

# Business Recommendation

- The company can advertise their cars to different groups of customers based on the results of our analysis. For instance, since age seems to correlate positively with price, they can highlight features of SUV that will be appealing to older people, or that features that would make younger people desire an SUV as they age.
- Most salary earners mostly prefer Hatchback. The company can set up schemes that makes it easier for more salary earners to buy less expensive cars.
- The company can introduce promos and features that can attract more females to buy expensive cars.
- There is no point spending resources to advertise cars to only people with certain educational level as analysis reveals that there is not much difference in the purchasing behaviour of customers across educational levels.
- Hatchback cars with more seats should be made available and advertised to people with dependents