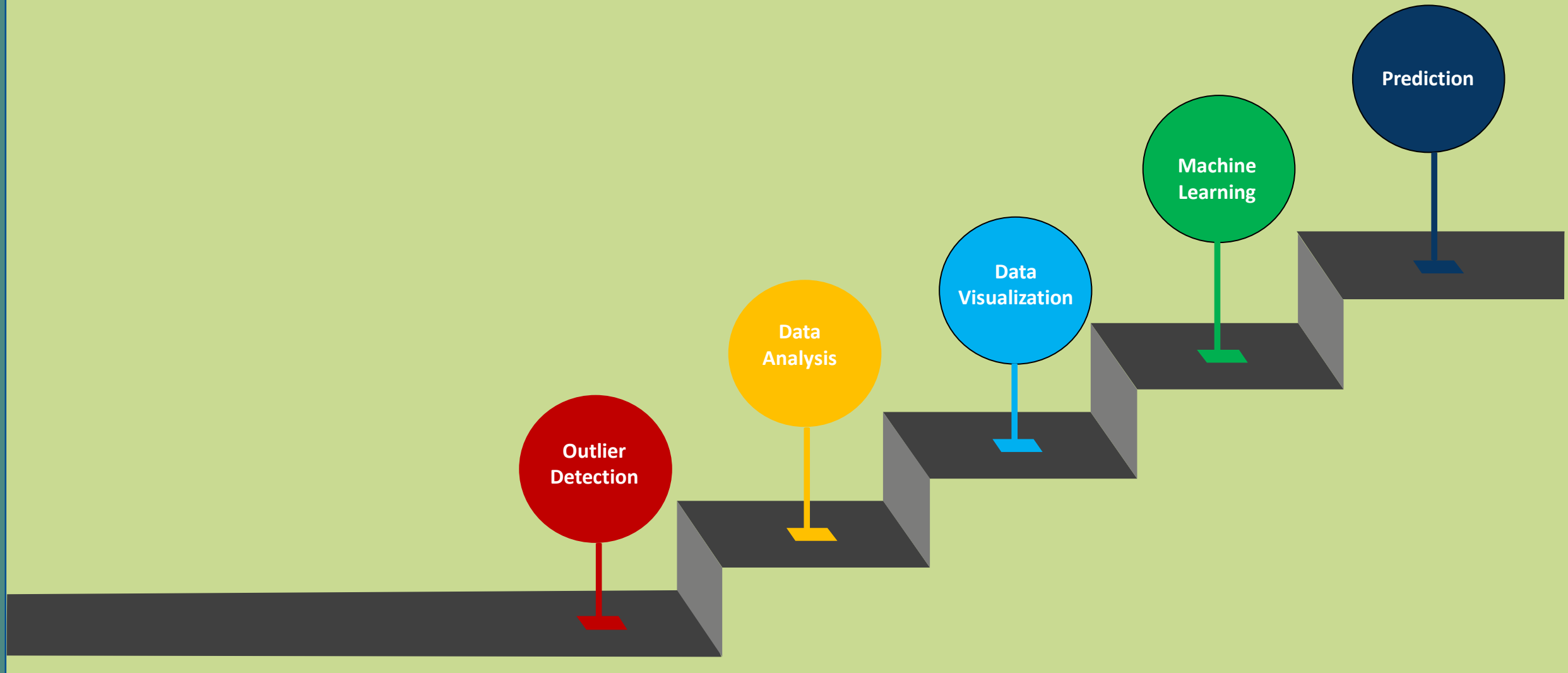
US CONSUMER TIME SPEND Analysis

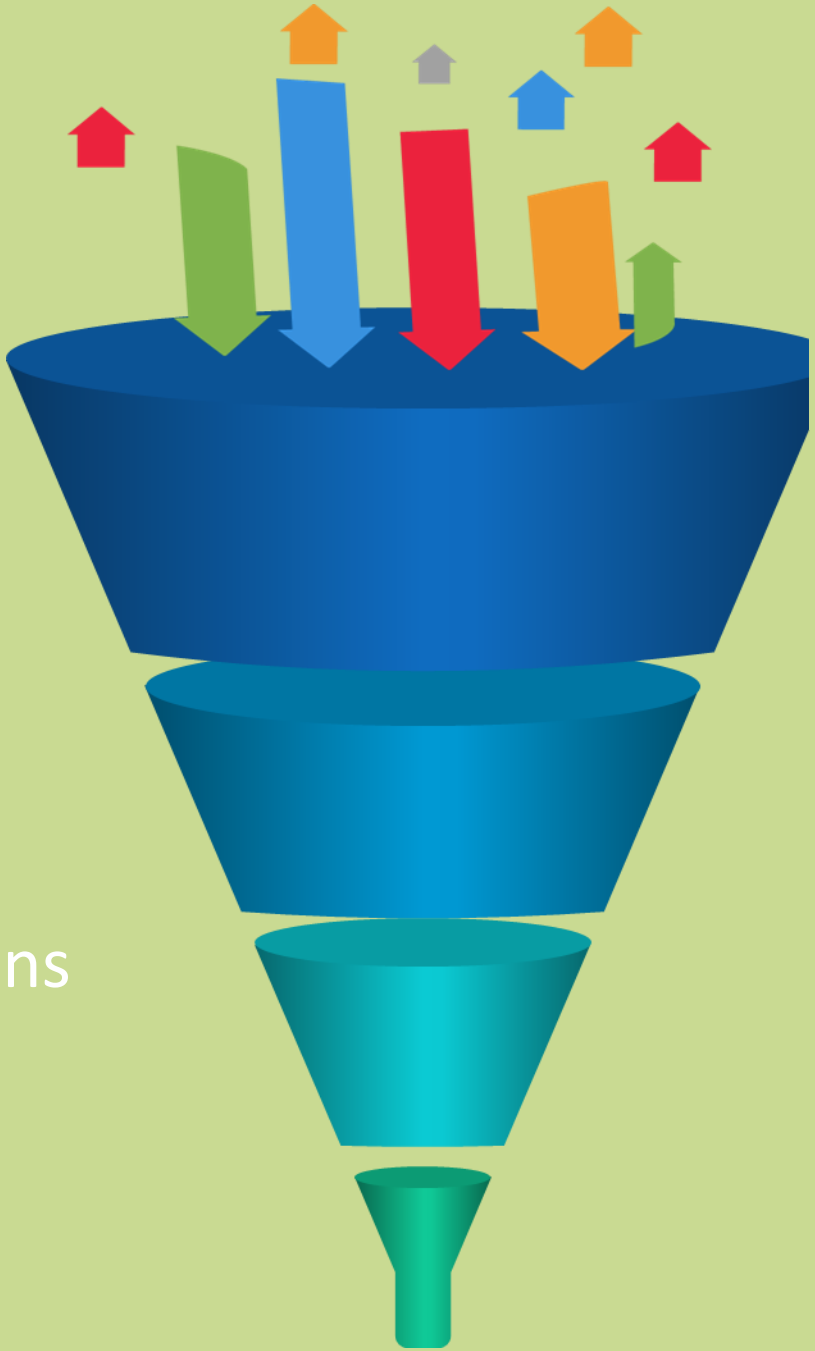
OBJECTIVE

- Survey based data collected between 2005 to 2012 used for consumer analysis
- Aim of the project is to
  - Identify change in the pattern when the great recession happened in 2008
  - Identify factors which impacts the shopping pattern of the consumers
  - Develop model to predict the employment status of an individual based on time spent by individuals on the daily activities

PROJECT OUTLINE



OUTLIER REMOVAL



The extreme observations that does not make any sense on average

Does it make sense for an individual to have no weekly hours but still earn??

Or..does it make sense for an employed individual to work but have no weekly earnings??

How about person who spends more than 14 hours of sleeping on average basis??

USING THE 98th PERCENTILE AS A THRESHOLD, REMOVED ALL OUTLIERS...

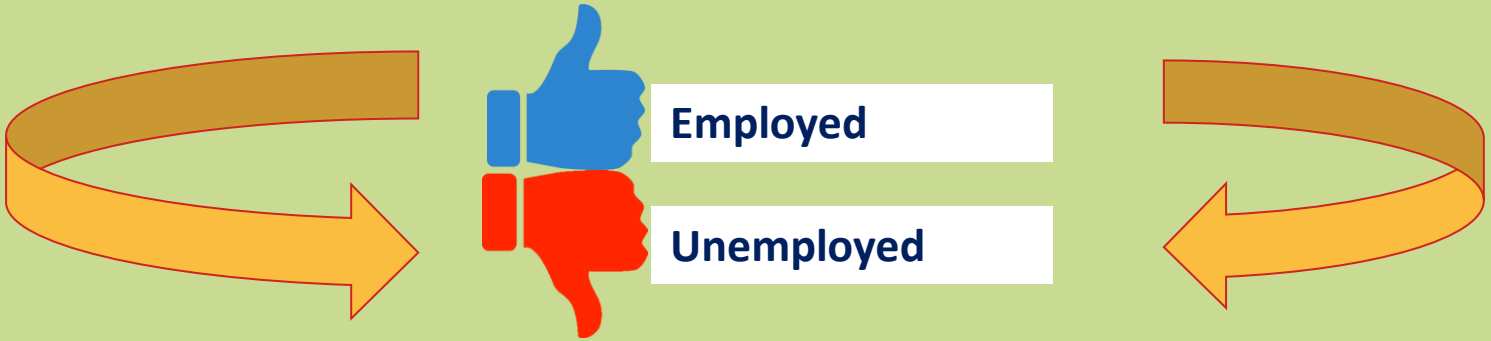
LEFT WITH 43429 OBSERVATIONS OF THE ORIGINAL 64000

General Findings

In 2012,

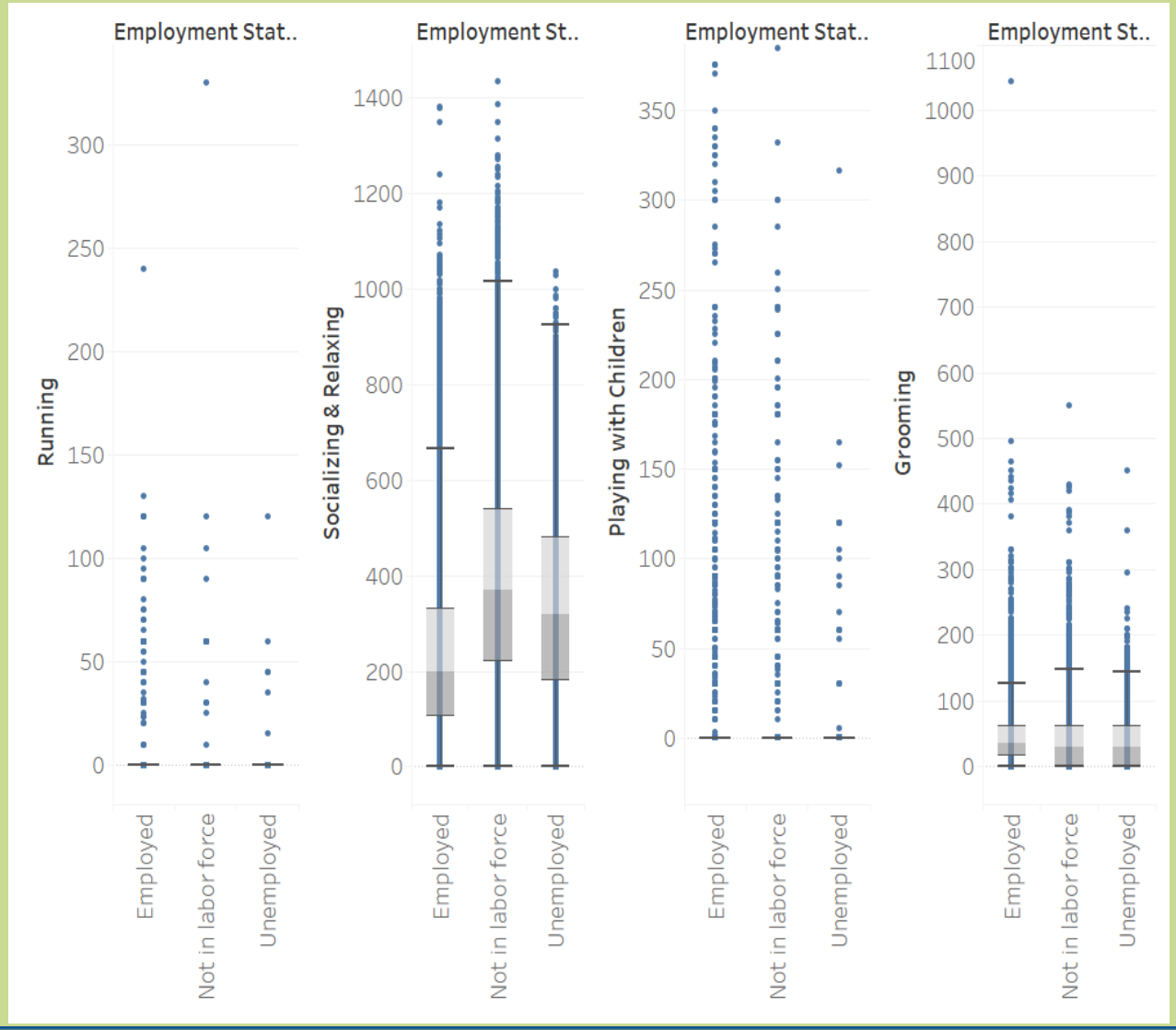
- On an average day, 88 percent of women and 68 percent of men spent some time doing household activities
- On an average day, 19 % of men did housework—such as cleaning or laundry—compared with 51 % of women
- On an average day, nearly everyone age 15 and over (95 percent) engaged in some sort of leisure activity. Men spent more time in these activities than did women (6 hours, compared with 4 hours)
- Adults living in households with children under age 6 spent an average of 2 hours per day providing primary childcare as compared to 45 minutes per day by adults with kids over age 6

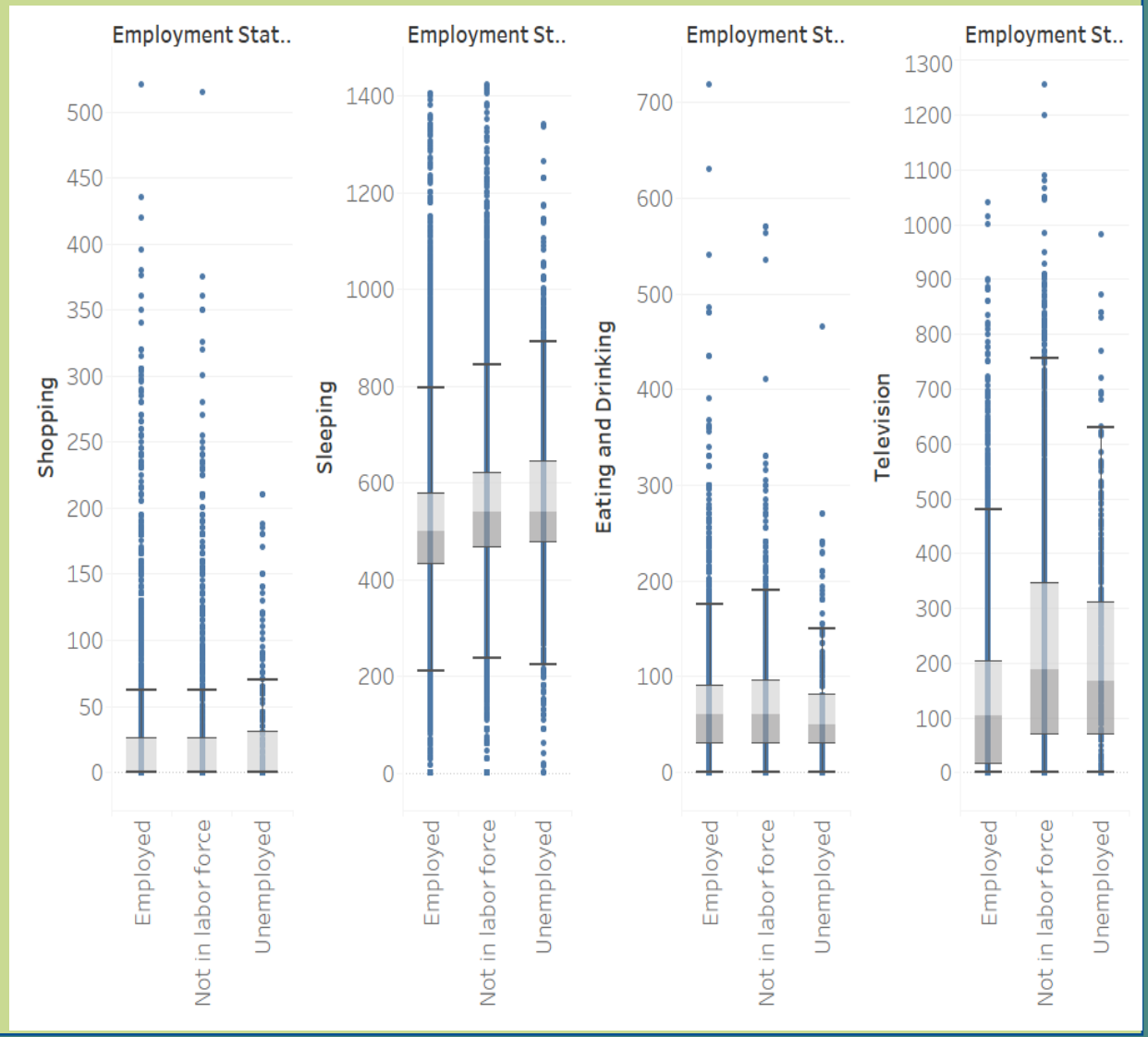
Summary of the time spending pattern in 2012 per activity



Here we are trying to understand how employment status influenced the time spending on various activities…??

Time Spend Pattern Across Activities





2012 Boxplots of Employment

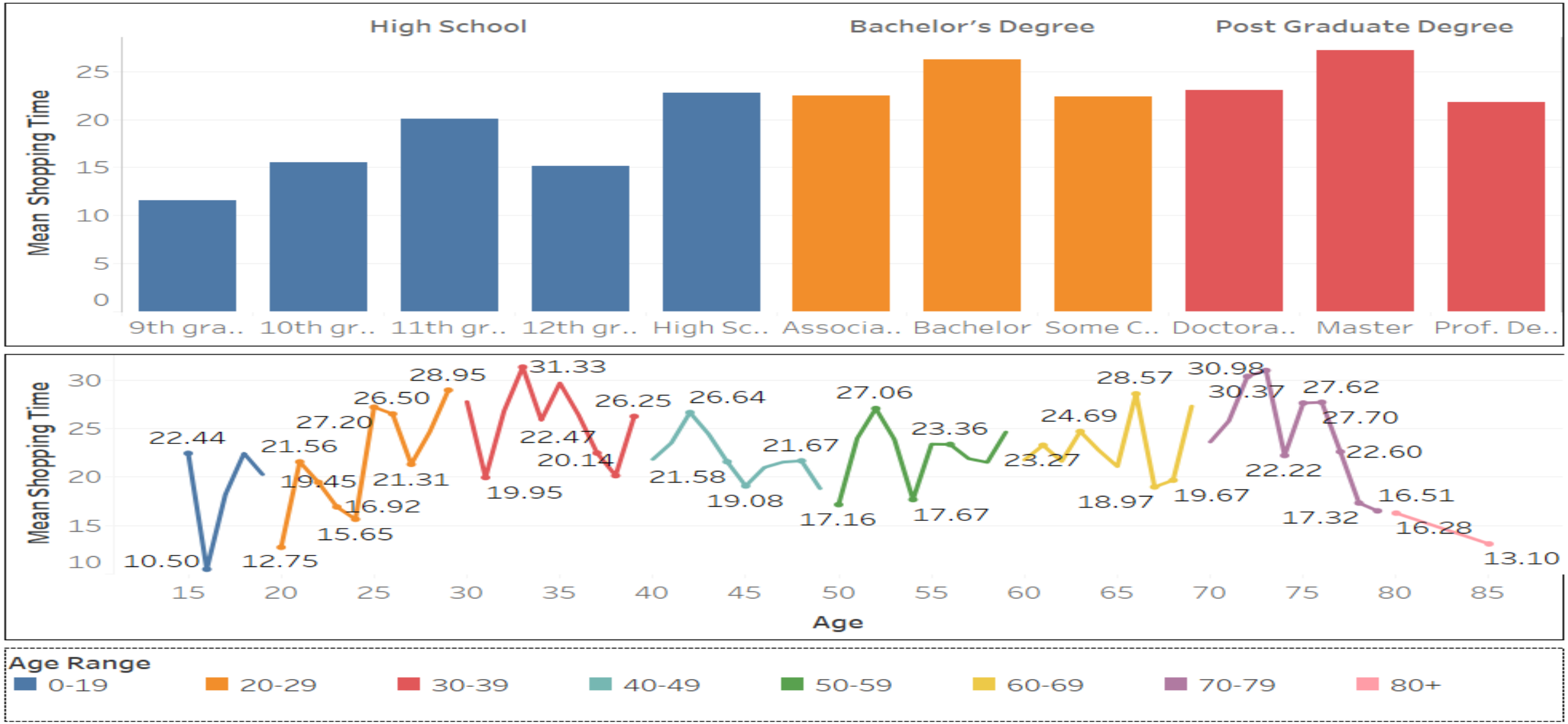
- From above visual inspection we can observe that means of each variables are different from each others

To confirm this ANOVA analysis is performed on time spent activities against Employment Status

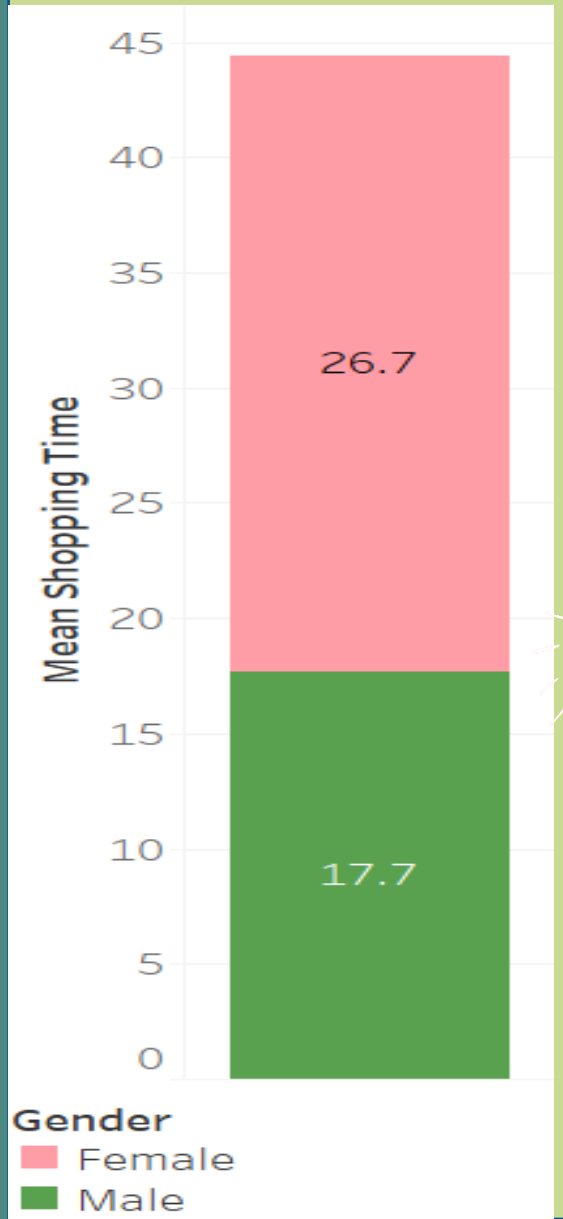
- Turns out means of all variables are significantly different from each other EXCEPT for SHOPPING Employment

The Average Shopping time is unaffected by employment status.

Does any factors impacts Shopping..??



Does any factors impacts the Shopping..??



Visually it can be analyzed that mean of the shopping pattern for Gender, Age and Education level is different.

To confirm this ANOVA analysis is performed on shopping time spent against Gender, Education Level and Age

Highly educated individual spent more time in shopping

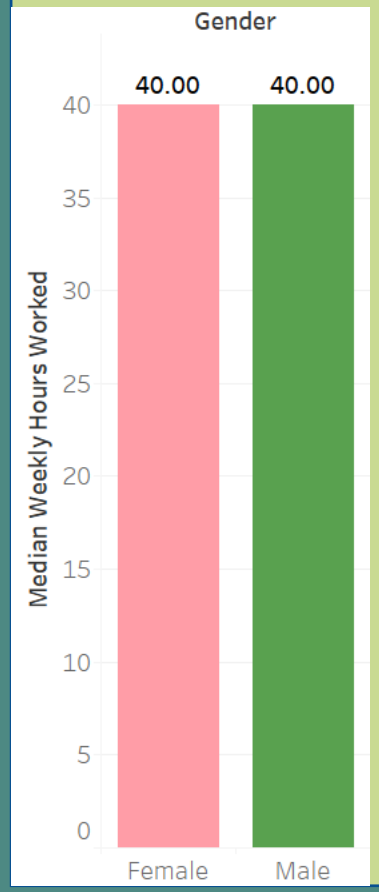
On an average female spent 50% more time on shopping than male

Individuals between age 30 to 60 spend more time on shopping compared to other age groups

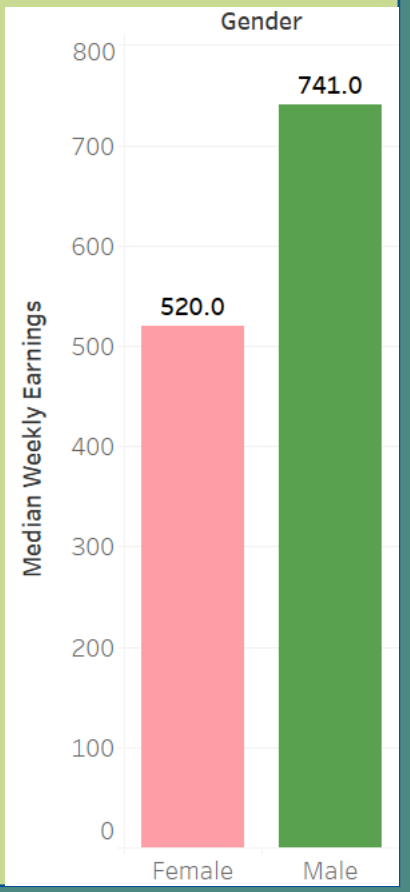
Is weekly earnings independent of Gender..??

As per the US labor bureau, a disparity exists between male and female earnings

To Confirm this ANOVA analysis is performed on Weekly earnings against gender



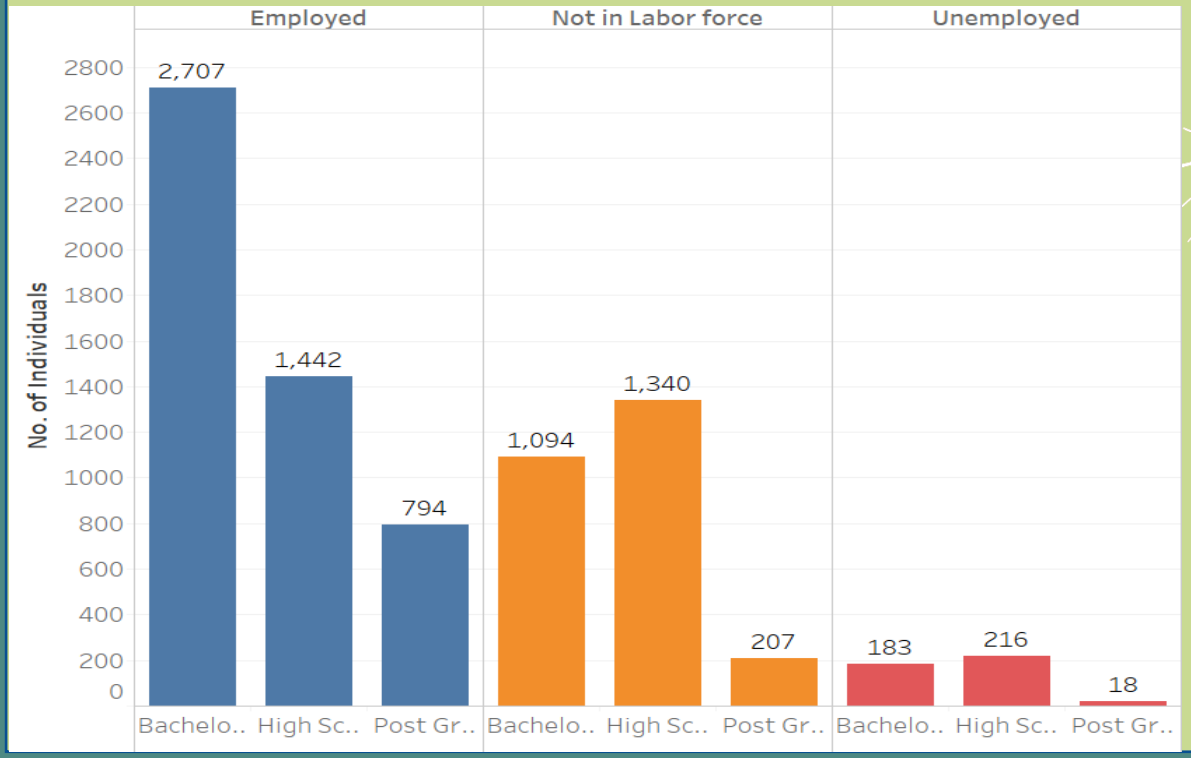
Despite of the fact that, male and female individuals work equally, male earn 42% more than the female counterparts



Is there any relationship between Employment Status and Education Level?

As per the economic theory, a higher education can help in getting job

To confirm this CHI-SQUARED analysis is performed on Employment Status against Gender and Education Level

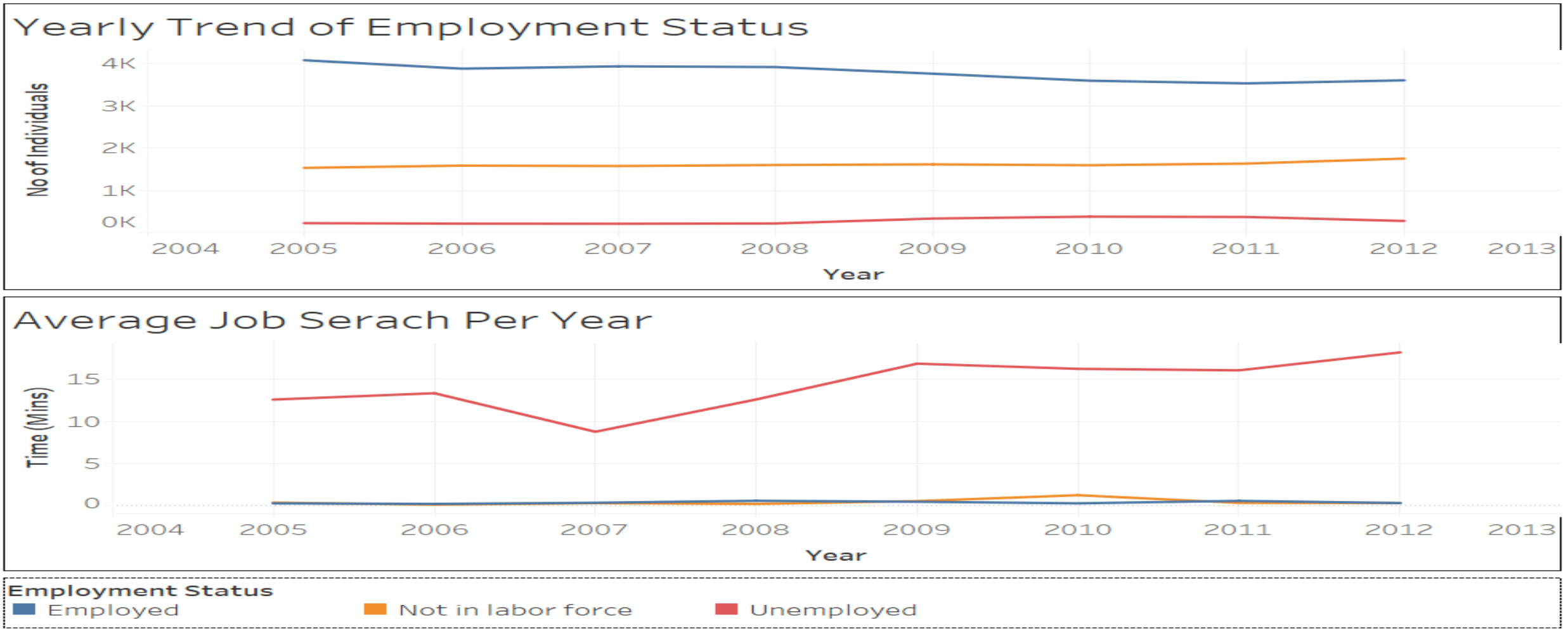


More than 60% of the employed individuals have at least bachelors degree

More than 50% of the individuals having at most high school education are not employeed

How Recession Affects the activities..?

We can observe that there is decrease in the employment status and increase in the job search activities between the year 2008 and 2010

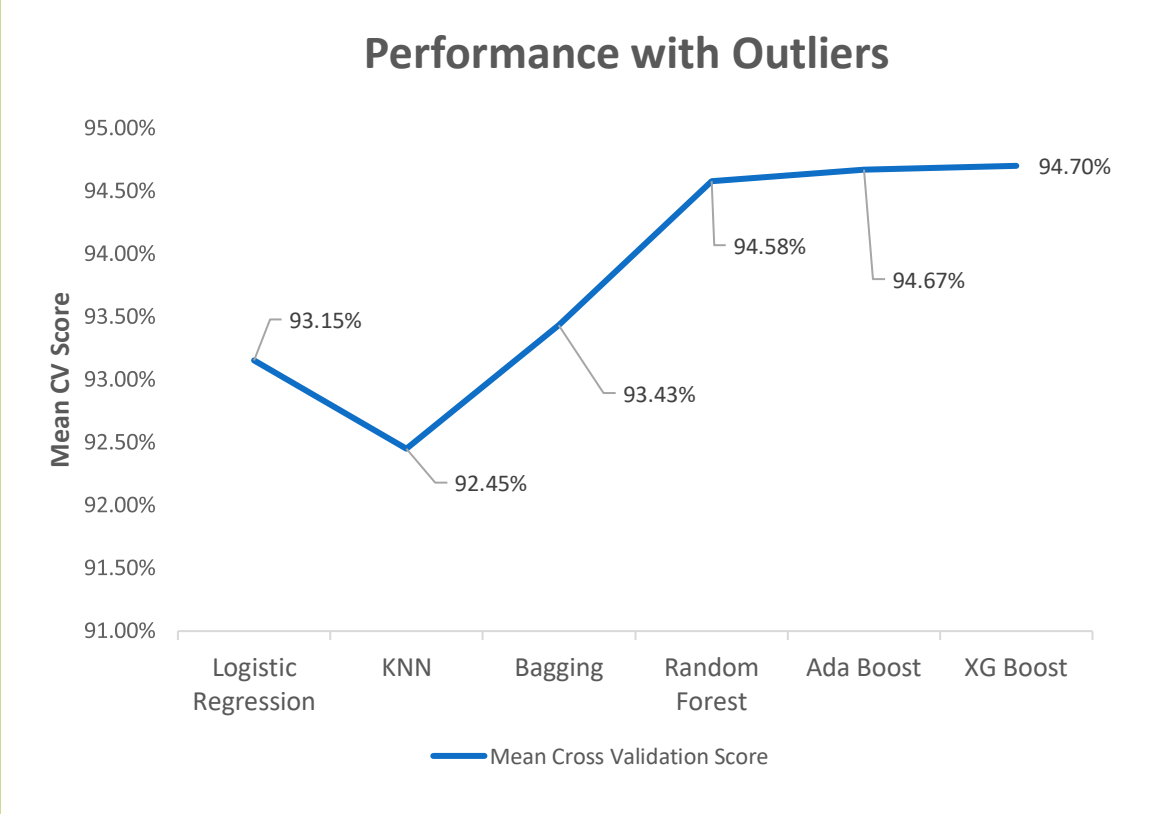


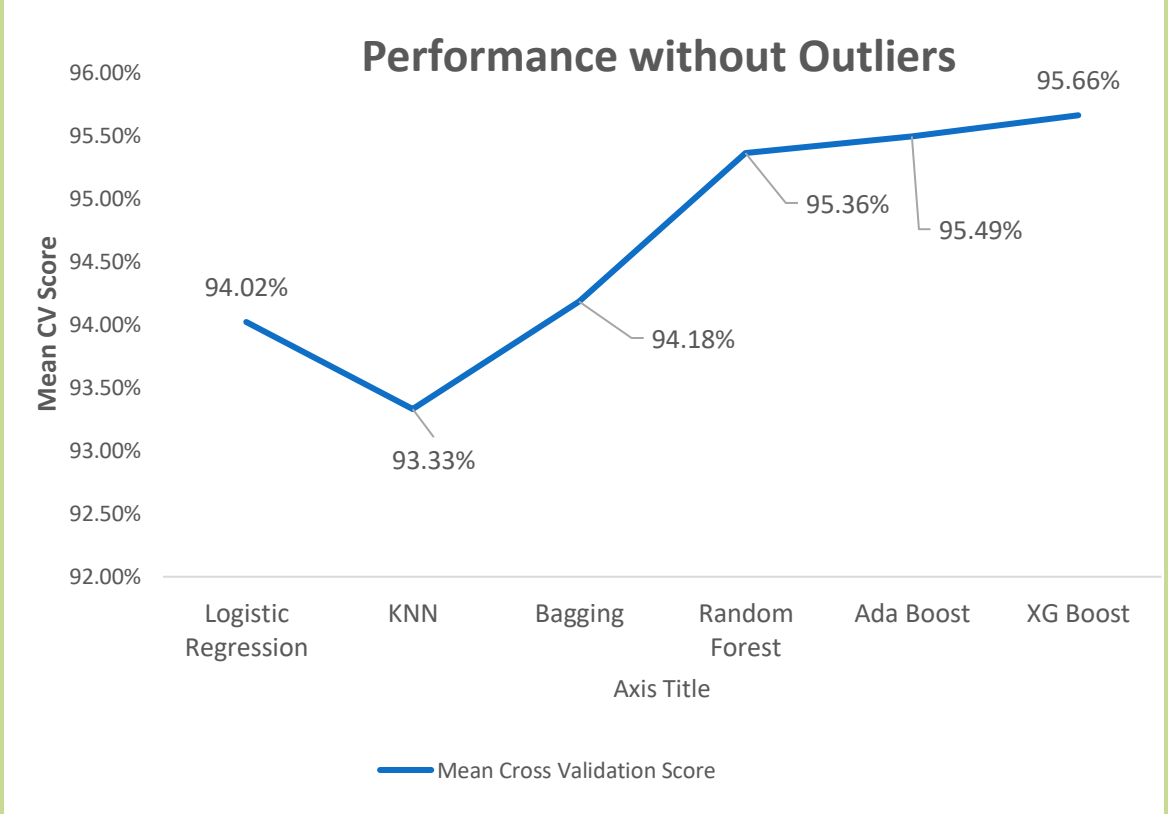
How Recession Affects the activities..?

There is a increase in the average time spend on the shopping by unemployed individuals as compared to employed.

However, we can see decrease in the daily average sleeping hours of the unemployed individuals compared to employed

Predicting Employment Status





Choose XG Boost Classifier over other models because, of high mean cross validation score and model performs better in terms of misclassifications (by almost 30% in False Positive)

THANK YOU!!!