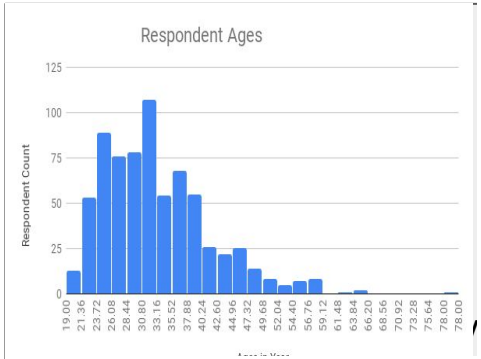
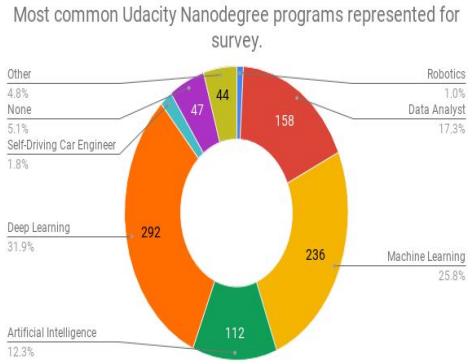
Project 2 : Analyze Survey Data

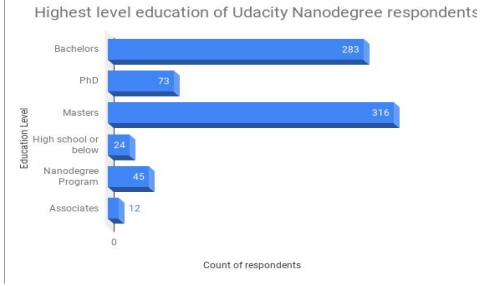
Some questions were attempted to answer from the survey date listed below:

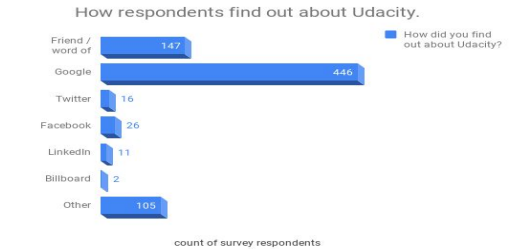
- What is the most common way people find out about Udacity?
- What is the highest level of education for most students and their minimum and maximum age?
- What is the most common Nanodegree program represented in the survey result?
- Do employed or unemployed correspondents spend more time to complete their Nanodegree Projects?
- Is there any relationship between spending time for learning materials and applying what learned?

1.Some Common Basic Insights from Survey data.









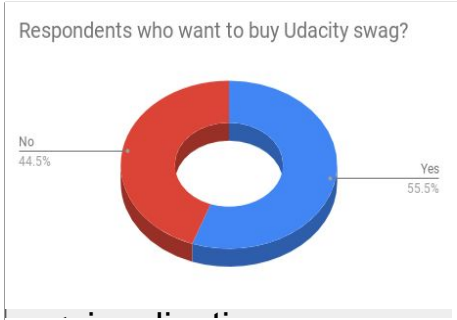
Some common basic insight selected and used to understand the survey data.

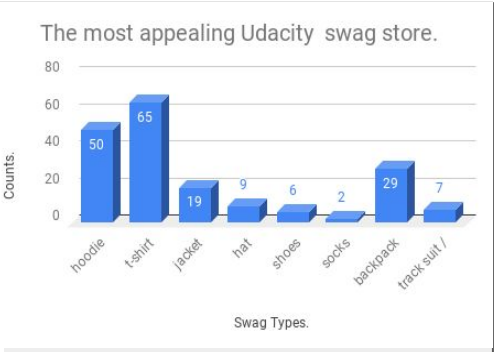
The maximum respondet age was 118 , the most common student's age which is the mode value is 29 and the average and the median was 35 and 32 years old.Therefore, the survey dataset has normal distribution property because they have closer similarity to mean and median.The reason for a slight difference between the mean and median is due to an outliers exist in the dataset.

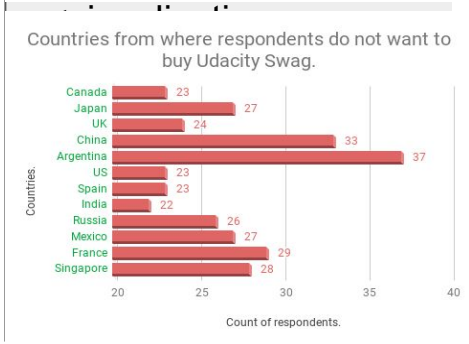
After removing some outliers the min and max ages are 19 and 78 and the mean and median are 33 and 32 respectively .The ranges is 59 and again the most common student's age is 29 years old aswell.

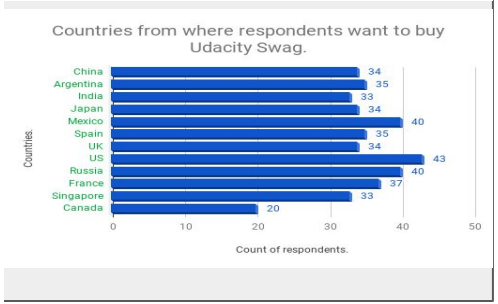
From the data the maximum highest education level was Master degree holders and the Google was used to find out about Udacity Nanodegree programs.The most Udacity Nanodegree programs represented for the survey was deep learning.

2.Udacity Swag Store.









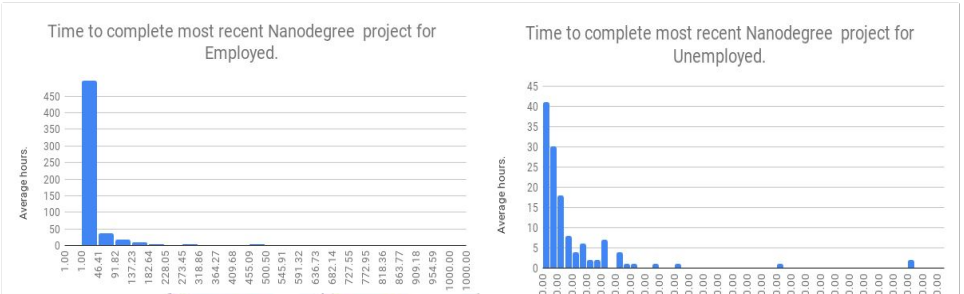
Pie chart,Histogram and Bar chart used to visualize Udacity Swag store.

From the pie chart about 56% of respondents were interested to purchase Udacity swag.

The most appealing swag was t-shirt.

From the dataset the maximum number of respondents from Argentina were not interested to buy Udacity swag while respondents from US was the highest figure who want to purchase .

3.Employed and Unemployed project completion time comparison.



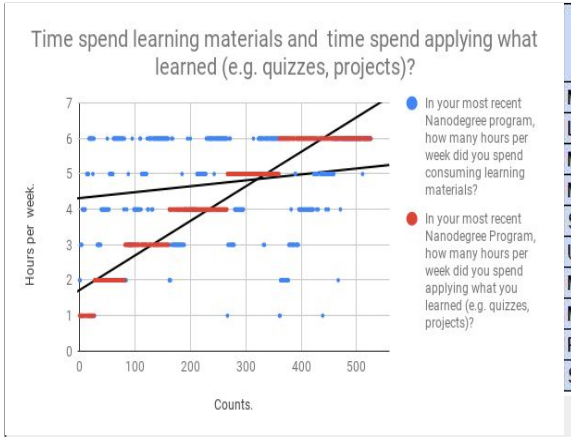
	Average project time for Employed(h)	Average project time for Unemployed(h)
Mean	32,888888889	37,0930232558
Standard Error	3,2861501605	6,28682409505
Mode	10	10
Median	12	15
First Quartile	7	7
Third Quartile	24	30
Variance	6220,0989372	5098,61627907
Standard Devia	78,867603851	71,4045956439
Kurtosis	63,405542468	28,1225852716
Skewness	6,982067518	4,92962754157
Range	999	499
Minimum	1	1
Maximum	1000	500
Sum	18944	4785
Count	576	129

Here are histograms and central measurements for employed and unemployed time spending to complete Nanodegree projects.

Both are right-skewed and contains an outliers and stretch the distribution far to the right.Their mean is higher than their median for both .From the distribution sample the mean is for unemployed 37 and for the employed 33 hours and the mode value for both datasets were 10 hours.Employed respondents spend the same maximum time as unemployed respondents to complete their Nanodegree projects.

The standard deviation for unemployed respondents was about 71 hours is way far from their mean 37 hours. Therefore ,unemployed respondents project time distribution have very large variable time spending.The reason might be due to many outlier present in the sample and the number of sample is smaller.Similarly for employed respondents it shows a higher variability spending time for a project to complete as well.In addition, the range for employed respondents is so wide and has an extreme observation value with some outliers.However,after removing the outliers both distributions have similar range value of 499. In contrast, there are higher observation sample values in employed than unemployed respondents.

4.Spending time of survey respondents learning materials and applying what was learnt.



	Time leaning materials(hour/ week)	Time applying what leard (hour/week)
Min	1	1
Lower Quartile	3	4
Mean	4.4	4.8
Median	5	5
Std	1.3	1.5
Upper Quartile	6	6
Max	6	6
Mod	6	6
Range	5	5
Sample	524	465

In this slide scatter plot and central tendency measurement table was presented for respondents from the survey spending time for learning materials and applying it.

From the scatter plot both spending and applying time have similar correlation. The reason is the trending line shows for both increasing .From the table their mean is for both 4.4 and 4.8 hours per a week respectively. Their standard deviation is also 1.3 and 1.5 hours respectively.Most of the respondents spending time is closer and similar to the average of the samples .From the mode values the maximum spending time for learning materials is the same as applying what was learnt hours per a week.

From the table the central tendency for both datasets (distribution) show the values are so closer.