



Data Foundations Nanodegree Program

Project#2 Analyze Survey Data

Name: Marwan Saeed Alsharabbi

Date: 02-11-2018

2018

INTRODUCTION

What is Project Analyze Survey Data?

In this project, you will analyze a real dataset about current Udacity students across a number of programs.

You will need to decide how to analyze the data and then communicate your findings about it. You will use spreadsheets to make your analysis easier. The goal is that you not only can perform an analysis, but also that you can communicate the results in a way that is clear and compelling.

Download the Analyze Survey Data table from the source.

What are the details of the Analyze Survey Data Table?

Why did you enrol in Udacity's Nanodegree programs?

Start a new career in this field
Grow skills for my current role

multiple choice

Help move from academia to industry
Help prepare for an advanced degree
General interest in the topic (personal growth and enrichment)
Other

What is your birthdate?

On average, how many hours of sleep do you get per night?

(if 0- home office)

What's your average daily commute (in minutes)?
On average, how many hours do you spend sitting per day?
On average, how many books do you read (or listen to) per year?
What city and state / province / country do you live in?
Do you want to buy Udacity swag?
Which item in the swag store appeals to you most?

Other.1

Which slogan / tagline appeals to you most?

Other.2

Are you employed?

What is your current primary occupation?

Other.3

Job Level

Other.4

What industry do you work in?

Other.5

How many years of experience do you have in your field of work?

What is the name of your current place of employment?

What is your highest level of education?

if they selected Nano degree see below

Intro to Programming

Business Analyst

Data Analyst
Machine Learning Engineer
Artificial Intelligence
Deep Learning Foundations
Self-Driving Car Engineer
Robotics
None

Other.6

What was most helpful when you got stuck in the Nanodegree program(s)?

Other.7

In your most recent Nanodegree program, how many hours per week did you spend consuming learning materials?

Other.8

In your most recent Nanodegree Program, how many hours per week did you spend applying what you learned (e.g. quizzes, projects)?

Other.9

On average, how many hours did it take you to complete a project in your most recent Nanodegree program?

What advice do you have for new or current students who aspire to complete a Nanodegree program?

How did you find out about Udacity?

Other.10

How likely is it that you would recommend Udacity to a friend or colleague?

What could Udacity do differently to improve your experience?

What are some additional subjects, courses, or tools and technologies that you'd like to learn at Udacity?

Is there anything else that you'd like to tell us?

Would you be willing to share more information for a chance to be featured on our blog, and help inspire others to be in demand and learn new skills with Udacity?

So it isn't perfect. It is a little messy (some things are input incorrectly, others are missing).

Cleaning Up the Data:–

All questions where the answer lives across multiple columns.

I changed some names of the columns in the table and hid some of them and deleted some after the merger.

Change the birthdate column to an age column by finding the number of days from that date to today, and then dividing by 365.

Some of the columns were used to analyze the Survey Data and implement the project.

I'm use spreadsheet application Microsoft Excel.

Please see attached Excel file.

I will begin the process of analyzing the Survey in the selected columns of the Survey Data.

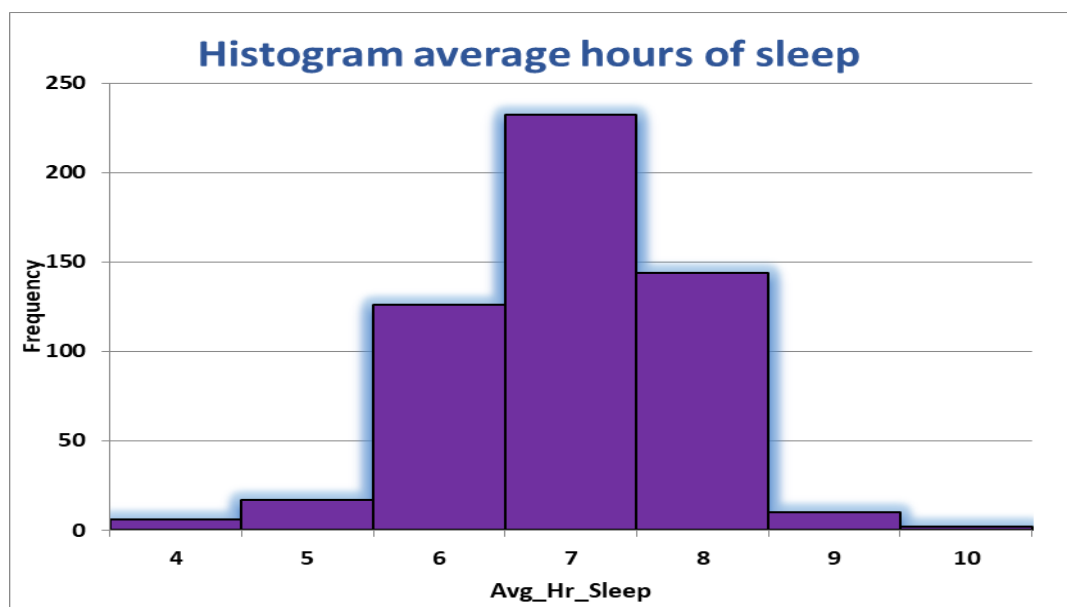
First column Avg_Hr_Sleep (by Histogram Chart)

Draw visualization showing the average number of hours of sleep by histogram?

- Delete the outlier data in the Avg_Hr_Sleep column
- Analyze the data in column and delete outlier values according to the methods set aside
- I'm delete the Outlier values that exceed the number of hours of the day because it is not normal and that sleeps one hour
- I'm calculate the mean, median, Max, Min, Q1, Q3 quarters, Rang, IQR and standard deviation
- I'm calculate Bins and Frequency (chart1 in excel attached)

Bin	Frequency
4	6
5	17
6	126
7	232
8	144
9	10
10	2
More	0

Describing quantitative	
Mean	6.93
Standard Error	0.04
Median	7.00
Mode	7.00
Standard Deviation	0.96
Sample Variance	0.93
Kurtosis	0.82
Skewness	-0.31
Range	6.00
Minimum	4.00
Maximum	10.00
Sum	5167.00
Count	746.00



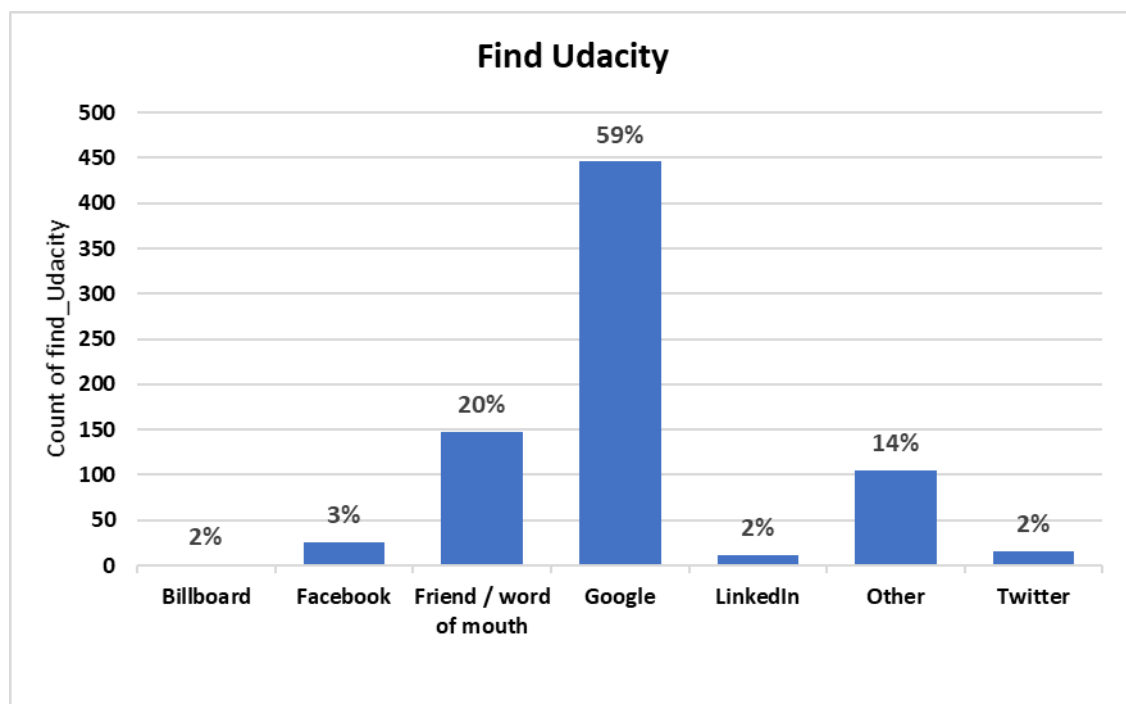
Through the visual I show most students sleep between 6 to 8 hours and the drawing shows that there is symmetric in the data and that Mean, Mode Median is equals This explains frequently normally distributed. This diagram represents the shape of the bell

Second column find_Udacity (by column chart)

What is the most common way people find out about Udacity?

- The column is taken from the original table
- Analyze the data in column replace empty cells with a word **other**
- I'm use Pivot Table and count the number of each method found in the Udacity of the social media or other (**chart4 in excel attached**)

find_Udacity	Count of find_Udacity
Billboard	2
Facebook	26
Friend / word of mouth	147
Google	446
LinkedIn	11
Other	105
Twitter	16
Grand Total	753



Through the survey analysis of Udacity students show how they got to this beautiful educational site that the Google site has The largest share as well as communication with friends, but we find that there are other means of communication did not mention in the Survey data was 14 percent may be other means

Third Relationship columns Age-years and Years_exp (by scatter chart)

Is there a relationship between age and years of experience in the Survey data?

Delete the outlier data in the Age column.

Replace the empty cells in the Years of Experience column by taking the average for the column.

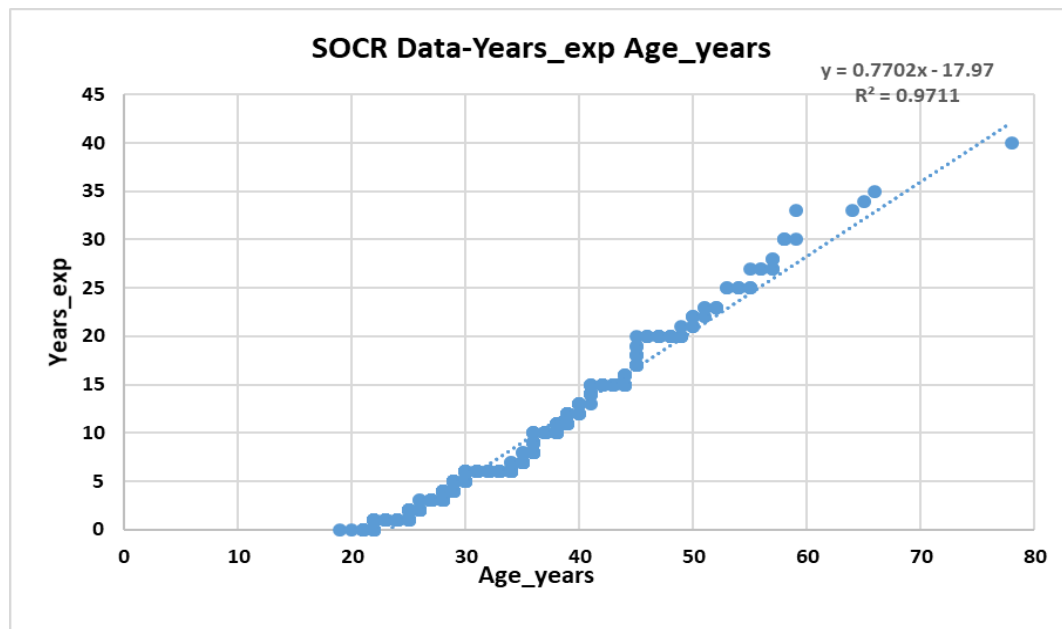
Deleting from years of experience values to come on bigger than old because he has illogical years of experience of greater age.

Rank the two columns of age and years of experience from smallest to largest.

Calculate measures of center and spread to describe quantitative data.

Age-years	
Mean	33
Standard Error	0
Median	32
Mode	31
Standard Deviation	8
Sample Variance	70
Kurtosis	2
Skewness	1
Range	59
Minimum	19
Maximum	78
Sum	23666
Count	712

Years_exp	
Mean	8
Standard Error	0
Median	6
Mode	6
Standard Deviation	7
Sample Variance	43
Kurtosis	3
Skewness	2
Range	40
Minimum	0
Maximum	40
Sum	5450
Count	712



Direction positively correlated, form liner, relationship strong noted through the Survey we note that the greater the age of the student increased experience with him and here a strong relationship between them (Chart5 in excel attached)

Fourthly relationship columns Study_hrs_wk and applying_learned (by Box Plot)

How many hours per week do students spend content and applying what you learned (e.g. quizzes, projects)? Does this differ by program?

Of the survey table project for Udacity took two columns (Study_hrs_wk, applying_learned).

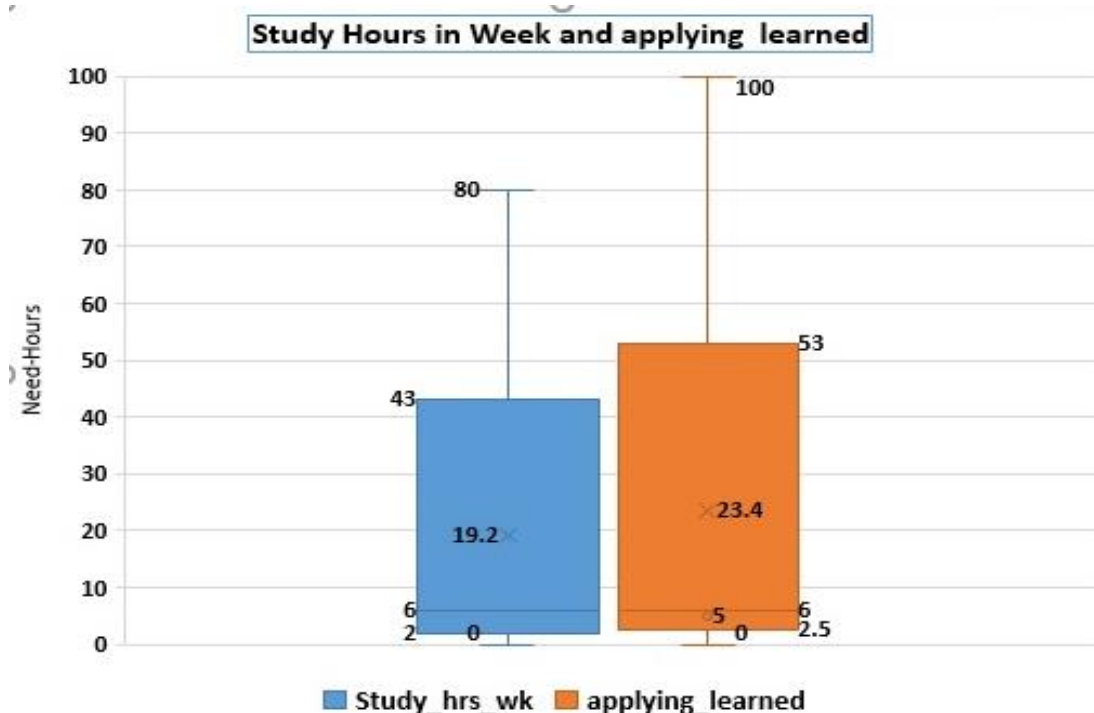
Rank the two columns of study_hrs_wk and applying_learned from smallest to largest. Reporting the 5 number summary values two columns.

Box Plot was drawn by reporting the values of 5 number summary for two columns

	Study_hrs_wk	applying_learned
Min	0	0
Q1	4	5
Median	6	6
Q3	6	6
Max	80	100

IQR	2	1
Upper Fence	9	7.5
Lower Fence	1	3.5
Rang	80	100
Mean	6.97	6.05
S.D.	6.54	6.42
Mode	6	6
count	747	747

Chart6 in excel attached



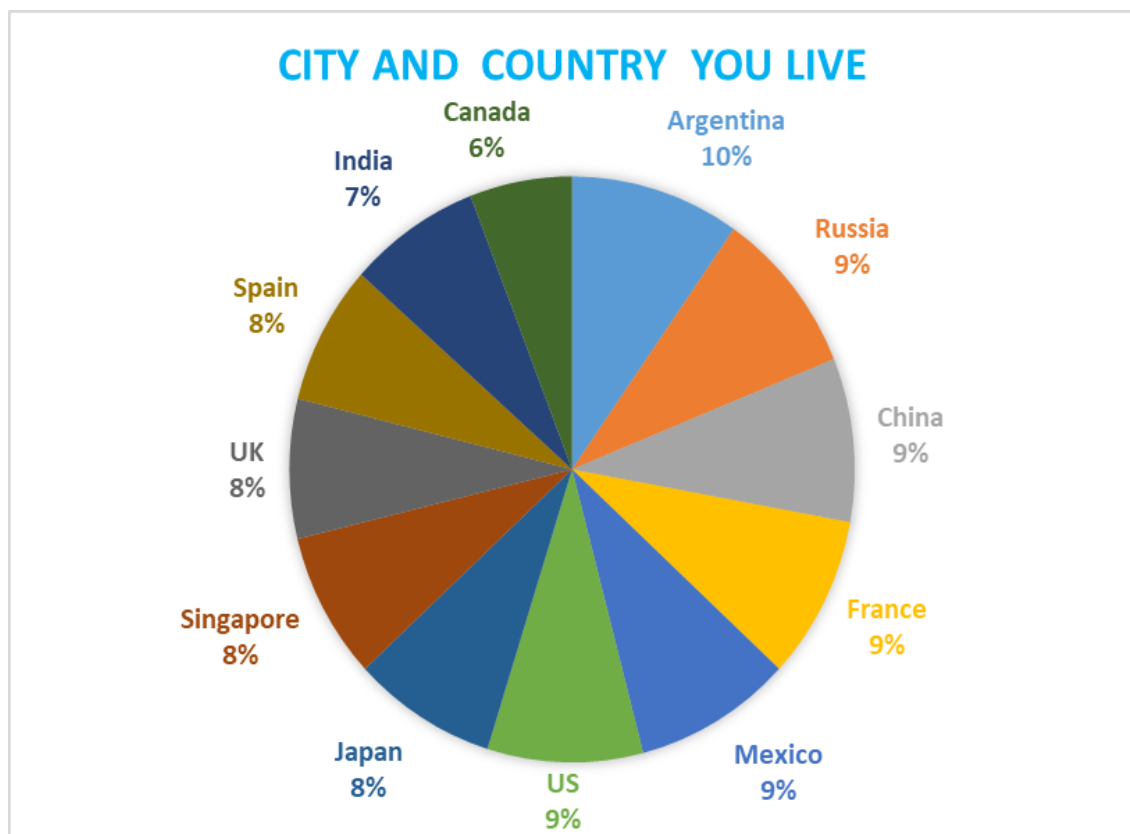
Noted in the visualization box plot is skewed to the right the mean greater the median positively skewed lots of the data observation to the left we can see now the distance between the median and the upper quartile is greater than median in the lower quartile

Fifthly column Country_The_live (by Pie Chart)

What are the most common countries/cities where students live?

I'm use Pivot Table and count the number of each method found in the Udacity of the countries/cities where students live (Chart4 in excel attached)

Country_The_live	Count of Country_The_live
Argentina	73
Russia	69
China	68
France	68
Mexico	68
US	67
Japan	62
Singapore	61
UK	58
Spain	58
India	57
Canada	44
Grand Total	753



Through visualization for the place where students live, the largest country came to Argentina, followed by Russia and Canada at the last level.

I wish success to all.