Bertelsmann-Udacity Data Foundations Nanodegree Program
2nd Data Science Project

Selma Yildirim

September 23, 2018

In this project, I decided to look at the data provided by the survey respondents with a Masters or PhD degree who stated that they are interested in their Udacity Nanodegree program to start a new career in that field.

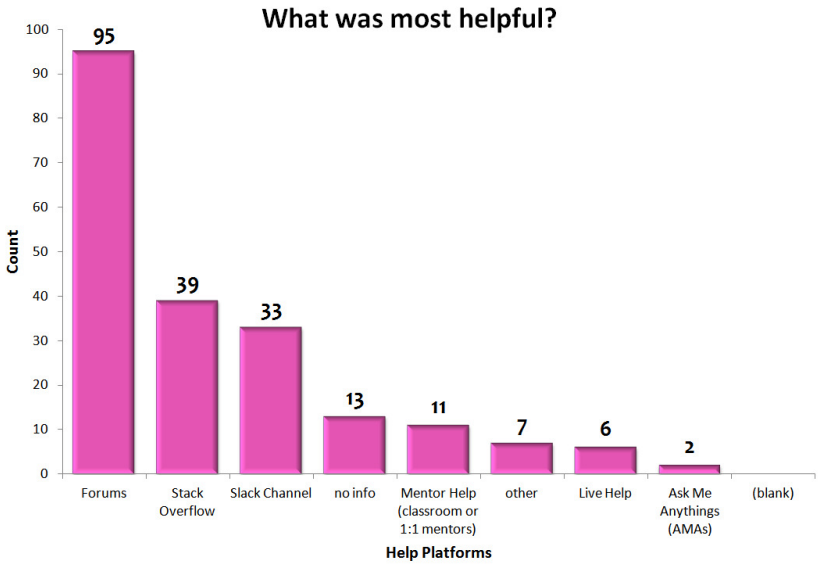
Data Cleanings Performed

- Columns that are not related to the research question and repetition of other columns are deleted.
- Data about people with Masters or PhD degrees and who want to start a new career in this field is filtered and a new worksheet containing this data is created.
- Columns are renamed in an understandable way.
- A new column combining columns related to what was most helpful in their learning is created by using two columns in the dataset.
- A new column for the number of Nanodegrees completed/enrolled is created.

What are the main characteristics?

- Out of 206 students who are interested in starting a new career, there are 175 with a Masters degree and 31 with a PhD degree.
- There are 57 students with a Masters degree and 10 students with a PhD degree who also wanted to grow their skills for their current role.
- Some students pursued more than one Nanodegree. On average, students with a Masters degree completed/enrolled in 1.23 Nanodegree programs while the average for the students with a PhD degree is 1.26.
- The most helpful platform that the students used when they got stuck was Forums followed by Stack Overflow and Slack Channel.

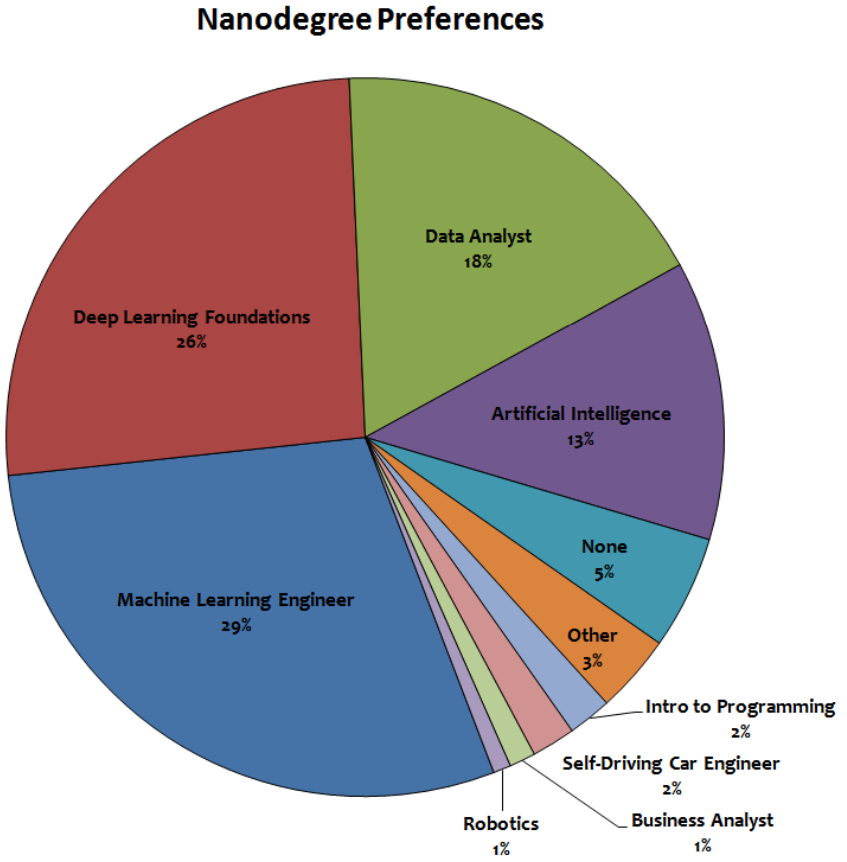
Education	Count	Count of grow_skills	Average of ND-count
Masters	175	57	1.23
PhD	31	10	1.26
Grand Total	206	67	1.23



What is the most popular Nanodegree?

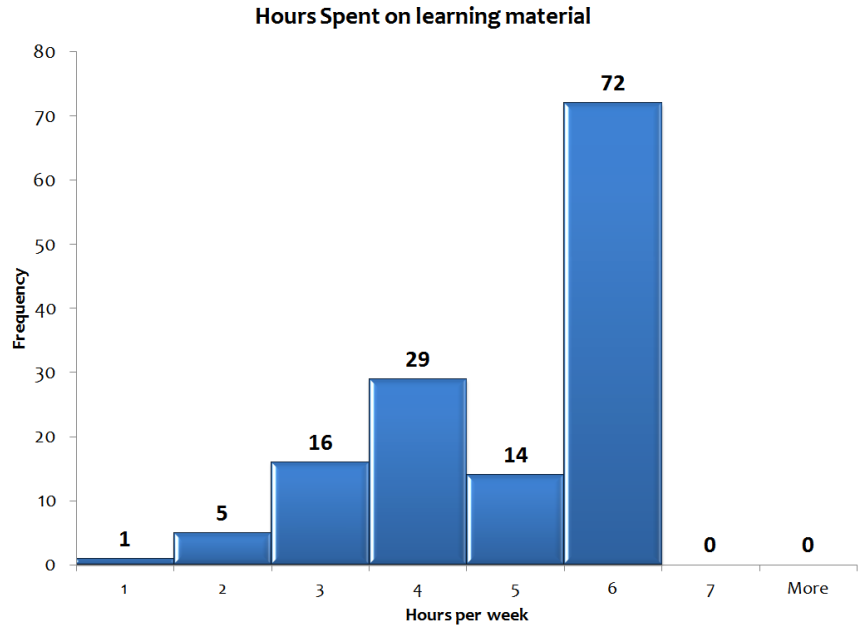
- The most popular Nanodegree is Machine Learning Engineer followed by Deep Learning Foundations.
- We should also keep in mind that some students pursue more than one Nanodegree (as it is shown in the table below). However, this fact does not affect the popularity of nanodegree programs.

Nanodegrees	1	2	3	4	Total
Count	169	27	9	1	206



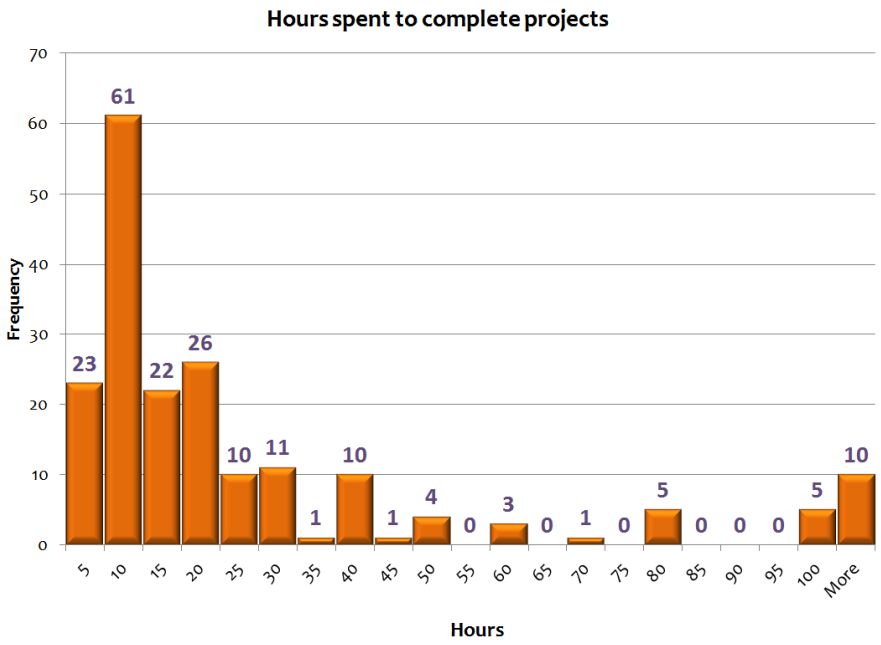
What are the learning study habits?

- The data for the hours spent per week on learning materials is left-skewed indicating that the mean is less than the median.
- Median, Mode, Q_3 and Max are all equal to the value of the tallest bar in the histogram. This means that at least 50% of students spent 6 hours on studying the material as this is the most repeatedly reported number.
- In addition, the statistics given in the table below shows that 50% of the students study between 4 and 6 hours.
- Standard deviation of the data is 1.29 indicating that the data is not varied a lot.
- The range is equal to 5, as the min and max number of hours reported are 1 and 6, respectively.



Mean	Median	Mode	Q1	Q3	IQR	Standard Deviation	Min	Max
4.94	6	6	4	6	2	1.29	1	6

How many hours do they spend on completing projects?



- Unlike the histogram about the learning habits, the data for the hours spent on projects is right skewed and it is more varied with range equal to 498 hours.
- In fact, each of histogram and a high standard deviation value confirms that the data is very spread out.
- In general, students spent a lot more time on projects than learning the material.
- Mean is higher than Q3=30 (i.e., 75% fall below this number). This would be due to the extreme outliers in the dataset. In addition, outliers effect mean more than median.
- Most frequently reported number of hours spent on projects is 8, which is shown by mode in the table on the left.

Mean	Median	Mode	Q1	Q3	IQR	Standard Deviation	Min	Max
36.38	15	8	8	30	22	78.34	2	500