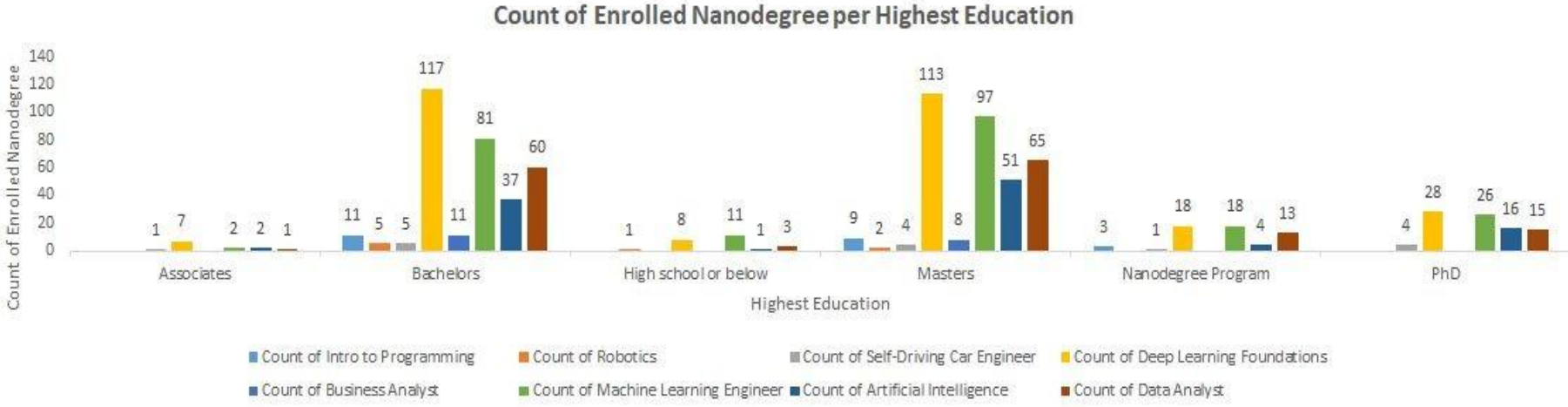
UDACITY DFND PROJECT 2

SUBMITTED BY

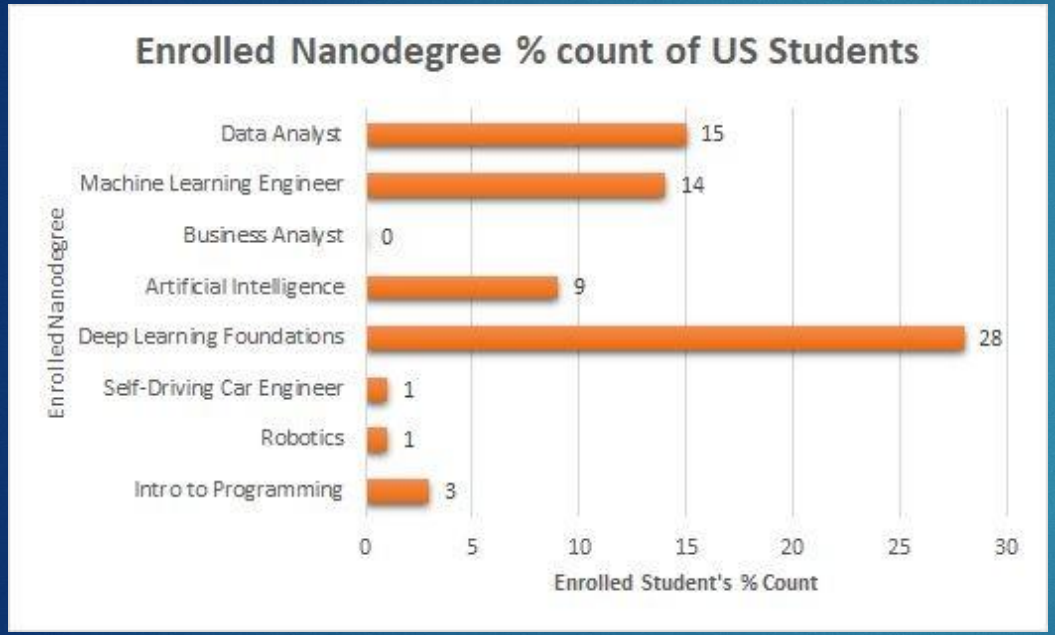
SIMRAN GUJRATI

What number of Masters students have been Enrolled in Data Analyst Nanodegree?



To answer questions such as **which program has students with what education background**, I have created the above **Bar Chart** with Education on X-axis and Count of each enrolled nanodegree of student on Y-axis. The Data Labels are shown on the top of each bar for the easy visualization. From the graph, we can say that Deep Learning Foundations is the most popular choice among all students except for students of high school or below who has shown more count for ML Engineer.

What % of US Students are enrolled in Robotics?



This is a **Bar chart** showing how many % of students from US have selected particular nanodegree which is created by filtering data of students from US and then having a count for each nanodegree.

We can analyse from the chart that, top 3 popular courses are DL foundations, Data Analyst and ML Engineer while students selecting for courses such as Business Analyst, robotics/ self-driving car is negligible.

Nanodegree	IP	Robotics	SDCE	DLF	AI	BA	MLE	DA
No of Students	3	1	1	28	9	0	14	15

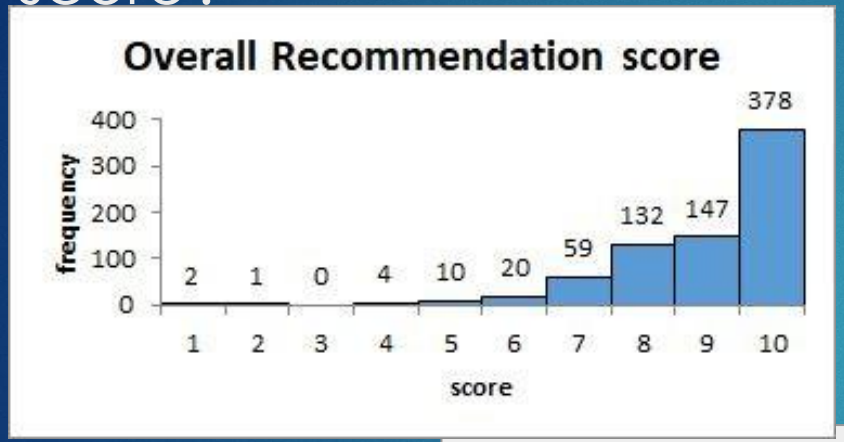
Does Average hour of sleep vary based on Employment?

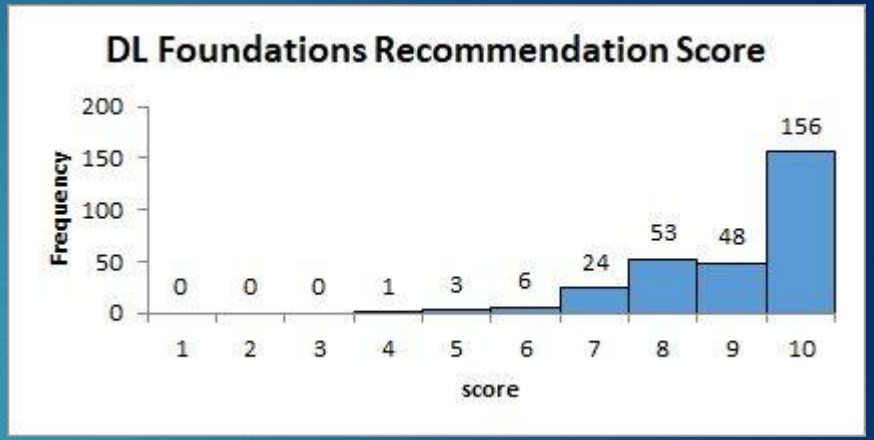
	Employed	Unemployed
Min	4	4
Median	7	7
Mode	7	7
SD	0.92	1.12
Max	10	10
Range	6	6

Comparing the employed to unemployed (after removing outliers from the data), **there does not appear much of a difference between the average number of hour they sleep per night.** Both of them follows gaussian/ normal distribution having similar minimum, maximum, mode and median (summary statistic).

While the **Mean** for employed candidate is 6.88 and for unemployed 7.13 hr/day which means unemployed candidate has much more sleep. **SD is greater for unemployed,** which means the variation in amount of sleep from mean varies higher in unemployed as compared to those who are employed. According to **Mode and Median**, all the candidates have same amount of sleep every night. Here, **Range** 6 which is difference between min and max, represents data is slightly spread from the mean and in total it is spread from value 4 to 10.

What is the relationship between overall recommendation score and DL Foundations score?





	Overall	DLF
Mean Score	8.98	9.07
Mode	10	10
Median	10	10
SD	1.36	1.22

The **average** recommendation score given by a student enrolled in any nanodegree **(after removing outliers from the data)** is 8.98 while the average recommendation score given by a student enrolled in Deep Learning Foundations is 9.07. That implies that not only DLF is more popular among the students but also they like the content and recommend to others.

Both the histograms are **left-skewed,** which should mean **mean < median < mode** but here, median=mode=10. We can say that mode is 10 means most student's recommendation score is 10. Value of **SD** is bigger for overall score as compare to DLF which infers variation of score from mean is larger for overall survey as compared to DLF.