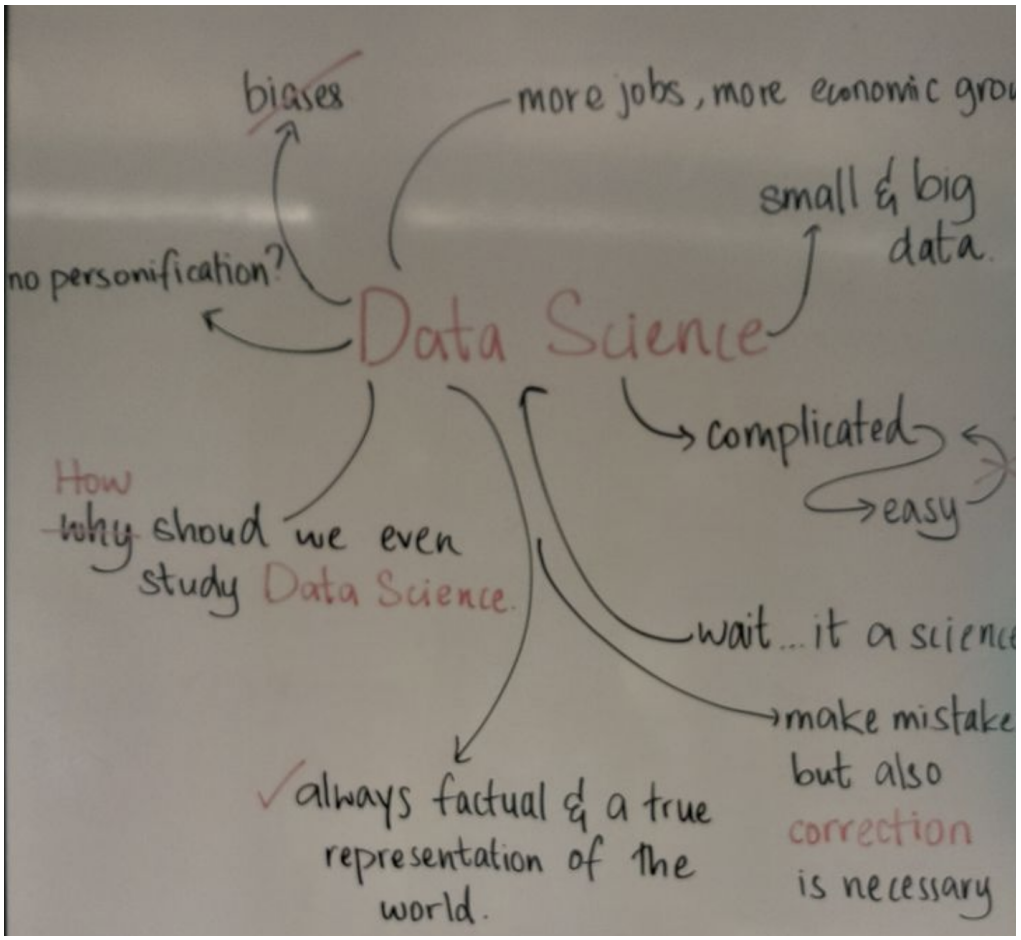
**Tenzin Uden**

Linked In | udent@whitman.edu | 509-859-7510| GitHub

---




@TenzinUden

---

Data Science is getting bigger and bigger and it's been used in anything you could imagine. From predicting the future to making bets and important decisions big or small.

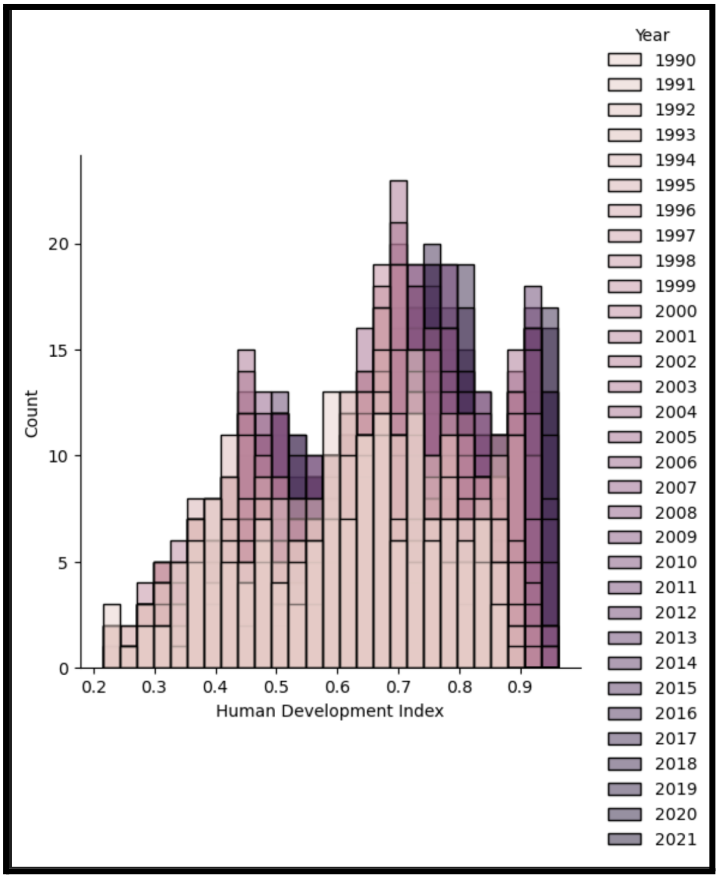
With change in time, using the datasets have also been hugely impactful in convincing world leaders to take actions against a lot of rising issues such as climate change, poverty, corruption and many more. While data science has its own benefits and good work that it does for the society, some data analysis could accelerate biases and wrong information that could impact the society at large. For example, when you do analysis on a dataset that is manipulated or spreads wrong information or doesn't have enough data to satisfy the requirements but the consumers only focus on the results, that could lead to misleading spread of information harming not just some but millions. Even if it was a factual representation of the dataset you were given, it still is not

accurate because of the fact that the datasets don't cover all aspects and don't hold enough value and information to legitimize the analysis.

"Of course when you ask people at Solitude where they like to ski, they will answer "Solitude." If they didn't like to ski at Solitude, they would be at Alta or Snowbird or Brighton instead. The superlatives he'd heard in praise of Solitude that day were not randomly sampled from among the community of US skiers.[1]"

A quote from Bergstrom & West, "Selection Bias," from Calling Bullshit, shows how asking certain and significant questions is important when you are collecting your datasets. We also need a specific goal and objective before we collect data so that we are not just wasting our time.

Thats why, being a Data Scientist could be very challenging as you will have to navigate through all of the aspects of a right data and also prevent the imposition of any basis that might alternate the datasets in a certain unpreferable direction.




---

"Complexity is an inherent feature of our existence—the world is rich in information that can be combined in endless ways.[2]"

Data is information and knowledge in the form of graphs, visualizations, figures and many more to make it easier for the

[1]Bergstrom, and West. "Selection Bias." *Calling Bullshit*, 23 February 2022, https://lms.hypothes.is/app/basic-lti-launch. Accessed 14 May 2023.
[2] Lupi, Giorgia. "Data Feminism." 23 February 2022, https://lms.hypothes.is/app/basic-lti-launch. Accessed 14 May 2023.

**Tenzin Uden**
Linked In | udent@whitman.edu | 509-859-7510| GitHub

people to understand facts and figures. Personally, data for me has been a good way or factor to predict future outcomes or use it as a resource to prove my findings.

*[1]Attached is an image of a data visualization on the Human Development index using data from Our World in Data conducted in project 5 of the Intro to Data Science .*

It's distinct from the information, knowledge and facts because data is a combination of all these in one as it contains all the information and the knowledge to turn into facts using figures. For instance, to get the data of your sleep schedule is to know the information about when you sleep, how often you sleep and the knowledge about your sleep schedule and what helps you sleep or disturbs your sleep. Then we can do a data collection and analysis on it to turn it into a fact that you could use to describe your sleep schedule.

To be a data scientist it means to know how to work around numbers and figures to answer small and big questions that can solve issues beyond the trade market, human safety, etc.

```
forbes_df["gender"].value_counts()
gend=forbes_df["gender"].value_counts()
gend.plot(kind='bar', x='x', y='y', figsize=(12, 8))
```

Also to be a data scientist we also have to be open minded and also consider all the factors that could make your data biased or incomplete.

*[2]Attached is an image of a formula and inbuilt functions used on created dataframes from the API data we collected on Forbes 2022 from Our World in Data conducted in project 9 of the Intro to Data Science .*

## SKILLS and ADVICES for a Data Scientist!!!

***Technical skills***: data wrangling, data visualization, web scraping, python programming language, logic building, research skills.

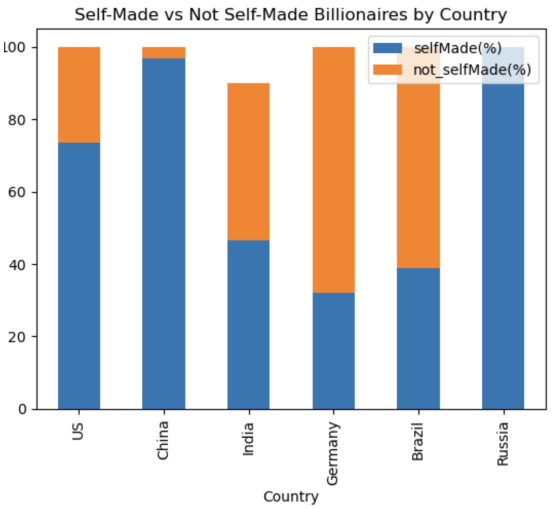
**Tenzin Uden**
Linked In | udent@whitman.edu | 509-859-7510| GitHub

Being able to break your data into its simplest form for the user to understand and also incorporate the message from it is one of the most important skills of a Data Scientist because we want to make people's lives easier.

Like in the image that you can see below, to get the data visualization to represent how much percent are self- made and how much are not, we had to do a whole calculation for different countries and also incorporate the formula to find these values.

*[3]Attached is an image of a data visualization to represent self-made and inherited billionaires from the API data we collected on Forbes 2022 from Our World in Data conducted in project 9 of the Intro to Data Science .*

```
c=forbes_df[forbes_df["country"]=="China"]

c_selfMade=c[c["selfMade"]==True]

c_selfMade.shape[0]/c.shape[0]

#Selfmade USA

usa=forbes_df[forbes_df["country"]=="United States"]
usa_selfMade=usa[usa["selfMade"]==True]
usa_selfMade.shape[0]/usa.shape[0]


# ## 73.529% of billionaires in the united states are self made

india=forbes_df[forbes_df["country"]=="India"]
india_selfmade=india[india["selfMade"]==True]


#Selfmade India
india_selfmade.shape[0]/india.shape[0]
i=india_selfmade.shape[0]/india.shape[0]
german=forbes_df[forbes_df["country"]=="Germany"]
```

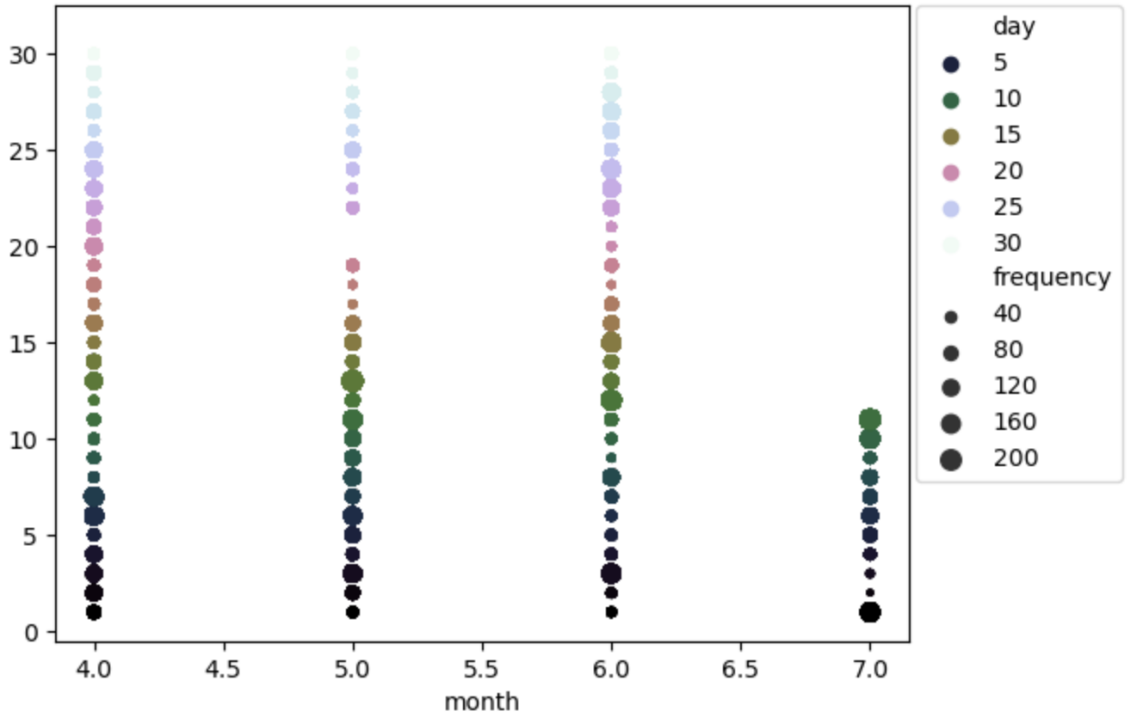



***Interpersonal skills***: critical thinking, organizational skills, presentational skills..

Datasets could range from millions to billions and more and knowing how to break them into chronicle order and also organizing it in a way that will be presentable to laymen is a very important skill.

*[4]Attached is an image of a data visualization created to represent monthly usage of Netflix and Spotify from two different users from project 8 of the Intro to Data Science .*

**Tenzin Uden**
Linked In | udent@whitman.edu | 509-859-7510| GitHub




---

"Failing to represent these limitations and nuances and blindly putting numbers in a chart is like reviewing a movie by analyzing the chemical properties of the cellulose on which the images were recorded. [3]"

Knowing your job as a data scientist is to provide right information in a constructive, and consumable manner that is able to provide not only a graph with numbers and lines on it but to tell a story of the interpretation of that dataset. An ***advice*** that I would like to give is to be ready to play with the datasets you are given and also make your logic or response to provide a readable dataset to laymen more efficiently and easier.

Be open minded and expand your horizons to not only the things you are taught in class but learning cool data wrangling stuff outside classes. Because why not when you have the resources.

---

[3] Lupi, Giorgia. "Data Feminism." 23 February 2022, https://lms.hypothes.is/app/basic-lti-launch. Accessed 14 May 2023.

**Tenzin Uden**
Linked In | udent@whitman.edu | 509-859-7510| GitHub

---

- *What kinds of problems can you solve?*

You can solve questions that involve the use of facts and figures because without those you wouldn't be able to create factual predictions or visualizations.

At large you can solve world problems by representing them in fractions and providing evidence of its impact on the world using the datasets you collected.

- *What kinds of questions can you not solve?*

You cannot solve questions that involve emotions because those are unpredictable and also maybe we cannot solve questions that have consistent data because then we wouldn't have anything to predict about or raise questions.

---

## Self Reflection

"No one really knows how big the Internet is, but some people say it's more than a "zettabyte," which, in case this means anything to you, is a trillion gigabytes or one sextillion bytes. That is a lot of brains in jars [4]"

The internet knows a lot of the things that a human brain doesn't know and that sparked my interest in drilling more into the world of Technology. Deciding to be a Computer Science Major and a Data Science Minor, I have come a long way since I first started to learn how to code. Everyday is a new learning and most of the time it's eye opening because you could do crazy things from developing a simple python game to scraping websites that reflect your interests.

My dream is to be a Major League Baseball Data Analyst and eventually work for FIFA as I am very passionate about sports and just imagining myself in those

[4] Lepore, Jill. "The Data Delusion." 23 February 2022, https://lms.hypothes.is/app/basic-lti-launch. Accessed 14 May 2023.

roles keeps me going and wanting to learn more about the world through Data and Technology.

Over the course of learning Data Science I have realized that Data doesn't have to be perfect, it just needs to be a true representation of your datasets. I used to think that your data should always tell something positive and have consistency in it, but after having worked with a lot of datasets, I have realized that it doest have to be a straight line or picture perfect graph. It's your responsibility to make it look good and for it to tell a story that is convincing and usable. I am constantly learning new things and it is widening my horizons which is a good exposure because it is preparing me to adapt to new changes when I actually start working for the world.

My ultimate goal as a Data Scientist is to make the world a better place through proper research and representation of the world issues and advocate to the people to take actions and also help make important changes in the way the society right now functions. Therefore I will keep learning new things and take data science to its full potential.

---

Note: Thank you for reading! Have a great day ahead.