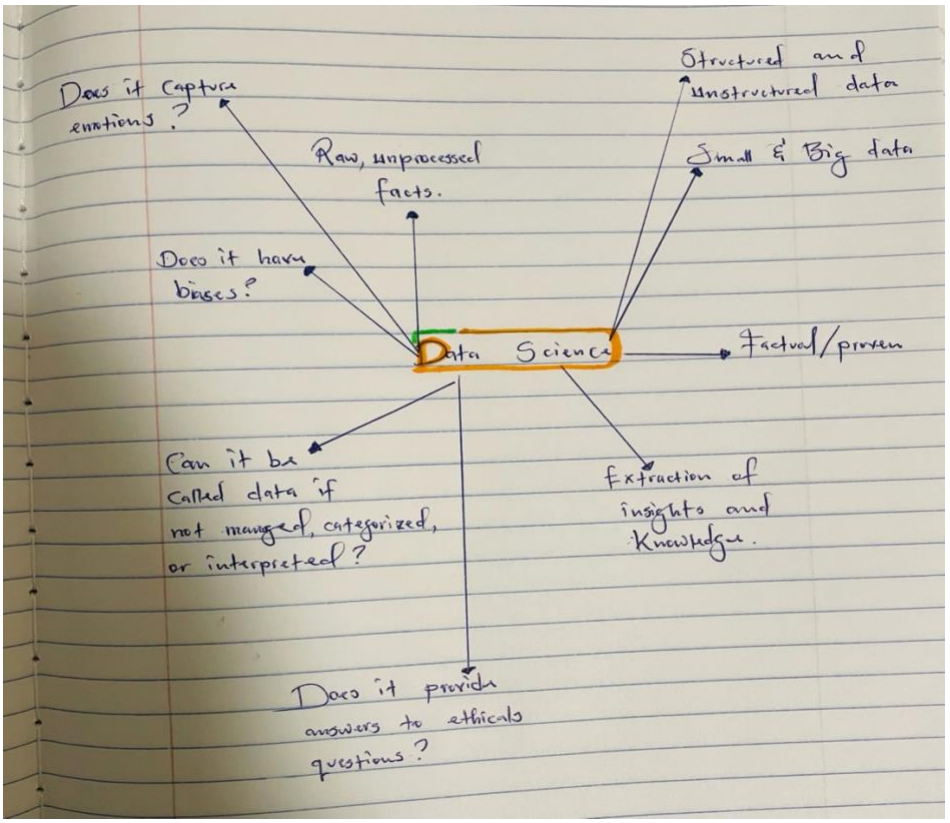
Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub



From Human-Centered Data Science

Data Science is expanding rapidly and finds application in diverse domains. Its capabilities range from predicting the future to facilitating decision-making, be it trivial or monumental. Over time, the utilization of datasets has played a pivotal role in persuading global leaders to address pressing issues like climate change, poverty, and corruption. While the positive societal impact of data science is undeniable, it is crucial to acknowledge that certain data analyses can perpetuate biases and disseminate incorrect information, adversely affecting society.

For example, conducting an analysis on a manipulated dataset or one with insufficient information may yield misleading results. Even if the analysis accurately represents the given dataset, it may not be truly accurate due to the inherent limitations of datasets, which may fail to cover all relevant aspects and lack the necessary depth to validate the analysis.

The quote from Bergstrom & West's "Selection Bias," featured in Calling Bullshit, emphasizes the importance of asking specific and significant questions when collecting datasets. Establishing clear goals and objectives before data collection is essential to avoid wasting time and resources. This complexity underscores the

Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub

challenges faced by Data Scientists, who must navigate through various considerations to ensure the accuracy and relevance of their analyses.

The role of a Data Scientist is demanding, requiring the ability to guide people through understanding facts and figures. The process involves careful consideration of potential biases and limitations in the data collection process. Having a specific goal and objective is crucial to ensure meaningful and impactful results.

As illustrated in the attached image of a data visualization on the World's internet usage, derived from Our World in Data in Project 5, data is distinct from information, knowledge, and facts. It amalgamates these elements, encapsulating information and knowledge to transform them into facts through numerical representations.

Data, therefore, serves as a comprehensive resource for predicting future outcomes and supporting findings. For instance, collecting data on one's sleep schedule involves gathering information about when and how often one sleeps, along with knowledge about factors influencing the sleep schedule. Through data collection and analysis, this information can be transformed into factual insights that describe an individual's sleep patterns.

Being a data scientist entails adeptly working with numbers and figures to address both minor and major queries, contributing to solutions that extend beyond the confines of a specific industry or trade Market human safety etc

```
# The GDI varies from time to time just because
Nigeria = hdi_df[hdi_df['Entity'] == "Nigeria"]
sns.scatterplot( x = "Year",
                 y = "Gender Development Index",
                 data = Nigeria,
                 hue = "Code")
```

Also, to be a scientist we have to be open-minded and consider all the factors that could make our 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 Gender Development Index from Our World in Data conducted in project 9 of the Intro to Data Science .

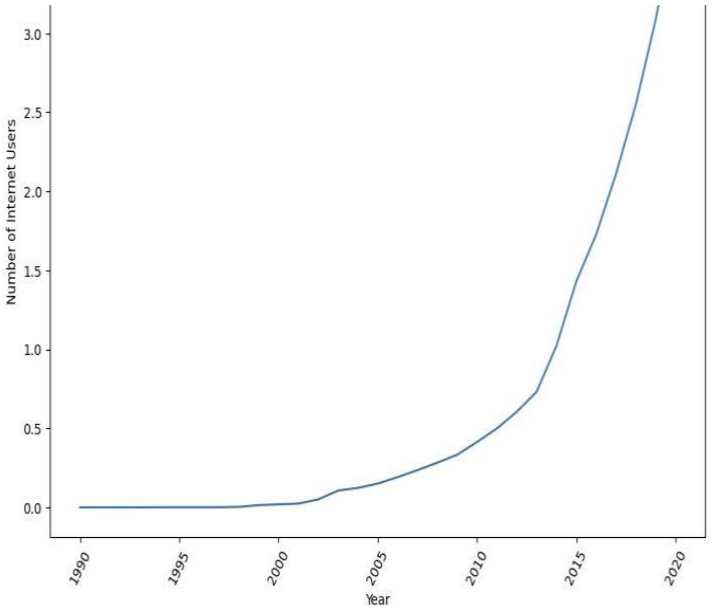
SKILLS and ADVICE for a Data Scientist

Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub

Technical skills: *Problem-solving/critical thinking,* data wrangling, data visualization, web scraping, python programming language, logic building, and research skills.

One of the crucial skills for a Data Scientist is the ability to simplify complex data for users and convey meaningful messages. The goal is to enhance people's lives by providing clear insights. As illustrated in the image below, achieving an accurate data visualization depicting the percentage of self-made individuals versus others required intricate calculations for various countries. It involved incorporating specific formulas to determine these values.

[3]Attached is an image of a data visualization to represent how the internet is being utilized worldwide I collected from Our World in Data conducted in project 5 of the Intro to Data Science.

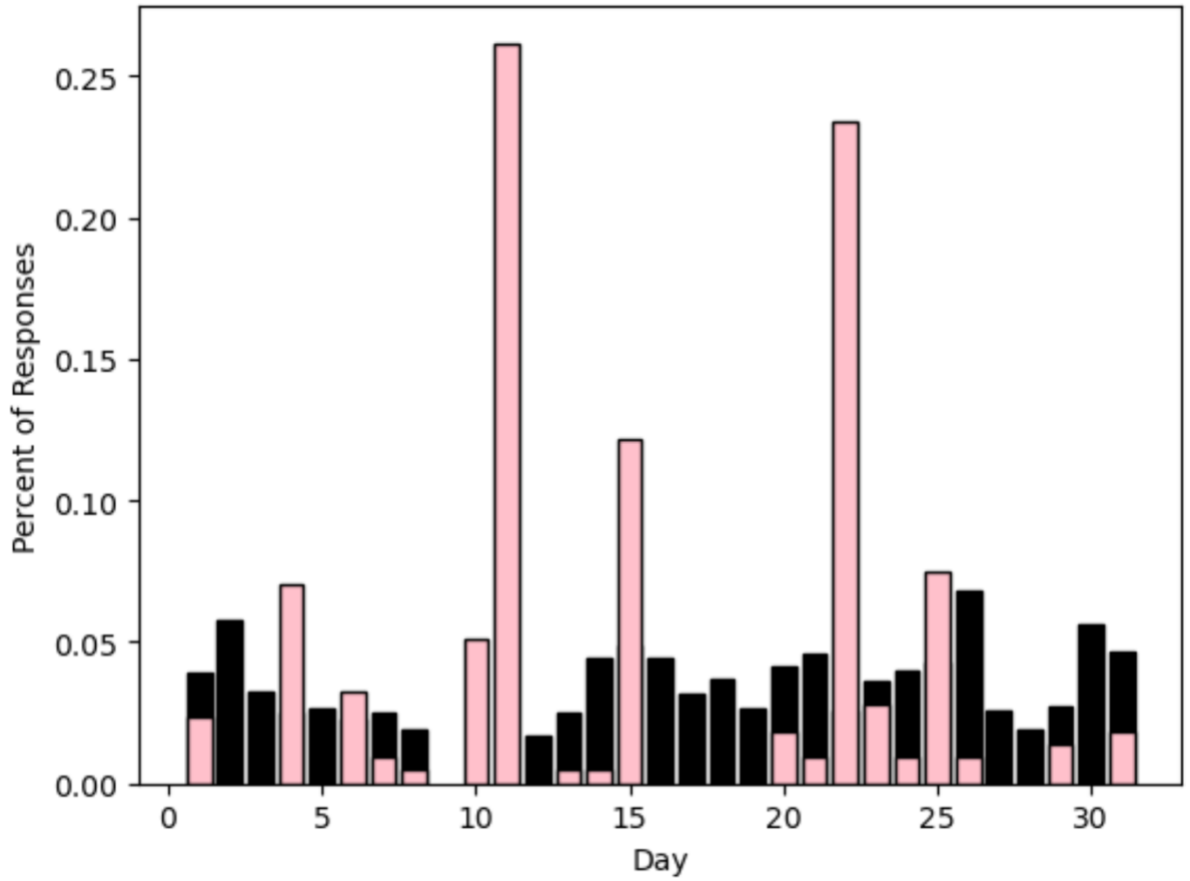


Interpersonal skills: critical thinking, organizational skills, presentational skills..

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

Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub

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



Failing to articulate these limitations and subtleties and mindlessly incorporating numbers into a chart is akin to critiquing a movie by dissecting the chemical properties of the cellulose on which the images were recorded.

Recognizing that a data scientist's role is to provide accurate information in a constructive and digestible manner, capable of delivering more than just a graph with numerical values and lines, is essential. My advice is to be adept at manipulating the datasets at your disposal and formulate logical responses that present data in a comprehensible manner to non-experts.

Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub

Maintaining an open mind and exploring beyond the confines of class teachings, delving into intriguing data-wrangling techniques outside formal education, is crucial. The availability of resources encourages this pursuit.

Problem-solving Capacity:

You can address questions involving facts and figures, as they form the basis for factual predictions and visualizations. On a broader scale, you can contribute to solving global issues by representing them quantitatively and offering evidence of their impact using collected datasets.

Limitations:

However, you cannot tackle questions rooted in emotions due to their unpredictability. Additionally, questions with consistently stable data may be unsolvable, as there is little to predict or interrogate.

Self-Reflection:

The Internet's vastness, often estimated at more than a zettabyte, fascinated me and led me to delve deeper into the world of technology. Pursuing a major in Computer Science, my journey has been marked by continuous learning and eye-opening experiences, from coding simple Python games to scraping websites aligning with my interests.

Throughout my Data Science journey, I've realized that data doesn't demand perfection; it requires authentic representation. While I initially believed data should consistently convey positivity and uniformity, working with diverse datasets has reshaped this perspective. Data doesn't have to form a perfect line; it's my responsibility to craft a compelling narrative.

Anthony Maniko
LinkedIn | manikoa@whitman.edu | 509-516-9511| GitHub

I aspire to leverage research and representation to make a positive impact on global issues. This involves encouraging action and instigating vital societal changes. I remain committed to continuous learning, eager to unlock the full potential of data science and contribute to creating a better world."