Assignment 6

Workbook link:
https://public.tableau.com/app/profile/oluwadunsin.agbolabori/viz/SalariesbyCollegetypeinUSAEduvizzers/Dashboard1

1. **Can you provide more context or specify the type of Tableau workbook to be replicated, such as data visualization, dashboard, etc?**

The name of the workbook is Salaries by college type. The dataset is from Kaggle and the design has been done by Oluwadunsin Agbolabori .The workbook shows various universities both public and private (including ivy leagues) and compares the starting and mid-career median salary of 5 of the following school categories :

- Ivy league
- Liberal arts
- Party
- State
- Engineering

It also shows the mid-career median salary for the 10^{th},25^{th},75^{th},90^{th} percentile.

This workbook has 7 visualisation/sheets and 1 dashboard. I have replicated all the sheets and the dashboard. All the annotations in all the sheets are in colour white therefore they are not visible in the worksheets as they are placed against a dark background in the original workbook's dashboard. I have done the same to replicate it.

2. **What criteria should be considered when choosing a workbook from the Tableau Public Gallery?**

Some of the criteria's that should be considered:

- How much the content of the workbook matches with an individual's interests ?

There are all kinds of dashboards on the Tableau Public Gallery and they cater to different people. What an individual wants to explore in and is easy to comprehend for them is what they should choose. From my experience , I downloaded many workbooks before deciding on the above as the other workbook did not cater to me. For e.g.: a Dungeons of dragons dashboard replication was hard to comprehend as I do not play the certain game so I as a student at an university choose the Salaries by college type dashboard as it caters to me.

- How accurate and detailed is it?

When doing data visualization ,it is important that how many aspects the visualizations are covering so a proper conclusion can be done for proper analysis. The more intricate the visualizations are, the more detailed the analysis will be.

- How much documentation is available?

Some of the Public dashboards only had images . Some of them did not have the data source downloadable. So for replicating a workbook ,the proper documents should be available or the links of the resources used should be mentioned so as to it can be accessed.

3.What role does data source play in the replication process of a Tableau workbook?
When trying to replicate a Tableau workbook , data source plays a major role as to replicate the same visualisation as filters might be used or some values might be omitted ,there could be changes made to the data or tables which is essential when replicating . Therefore ,if the correct data source is not used the visualization will not be same/accurate.

4.What impact does the level of complexity of the original workbook have on the replication process?
The level of complexity in this dashboard was medium. It had a prefixed dark background design with text embedded in it. So I used tableau tools to replicate the visualizations and the embedded text using floating text tiles. The level of complexity also decides the time and research needed for successful replication. It resulted in an incredible learning experience. As I learnt more intricate detailing in tableau and how to make it more visually appealing for better analysis.

5.What is the significance of documenting the replication process and what should be included in the documentation?

Documentation is required for replications to know all the intricate detailing and calculations for the correct visualization and also for using the correct filters, sets, parameters and calculated fields so that the data can be visualized accurately.

6.How will the new knowledge gained from replicating the Tableau workbook contribute to future data analysis and visualization projects?

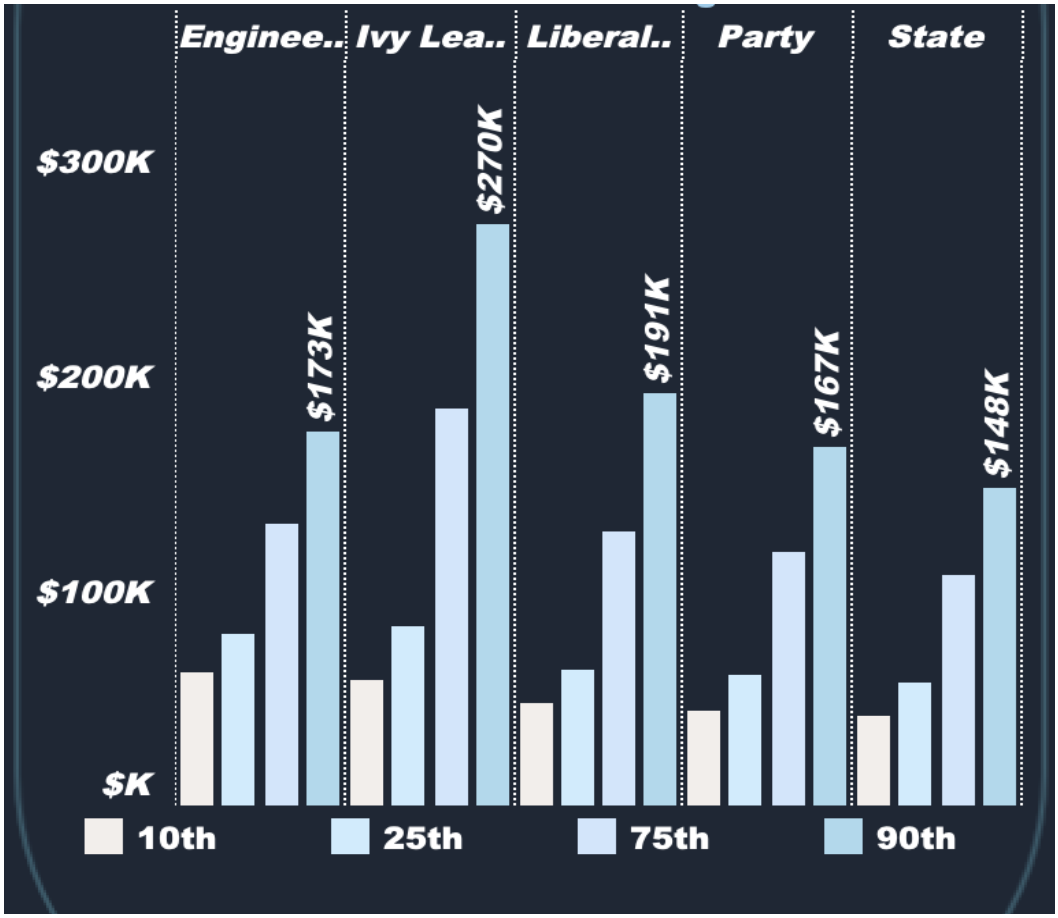
Replication of a public Tableau workbook has helped me gain knowledge of the more intricate features of tableau and how to use those to make more detailed and interactive visualizations resulting in more accurate and concise data representations.

7.What challenges might arise in the replication process and how can they be addressed?

The replicating of the intricate features ,assembling the dashboard and formatting were somewhat tricky but with time and proper research it was overcame. Spending time on researching the original workbook to be replicated, reading modules and reading tableau documentation can help address these problem.

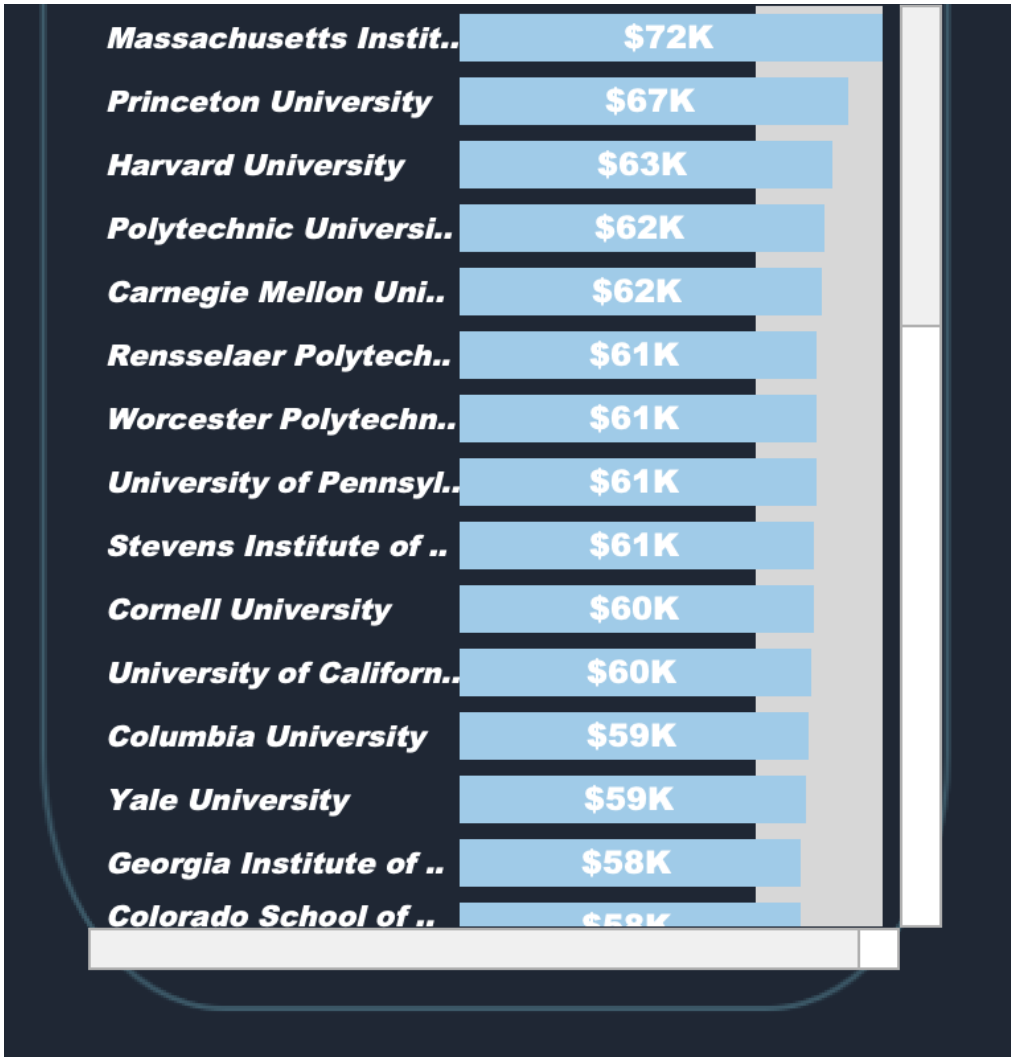
Visualizations done:

Mid-career median percentile vs School type



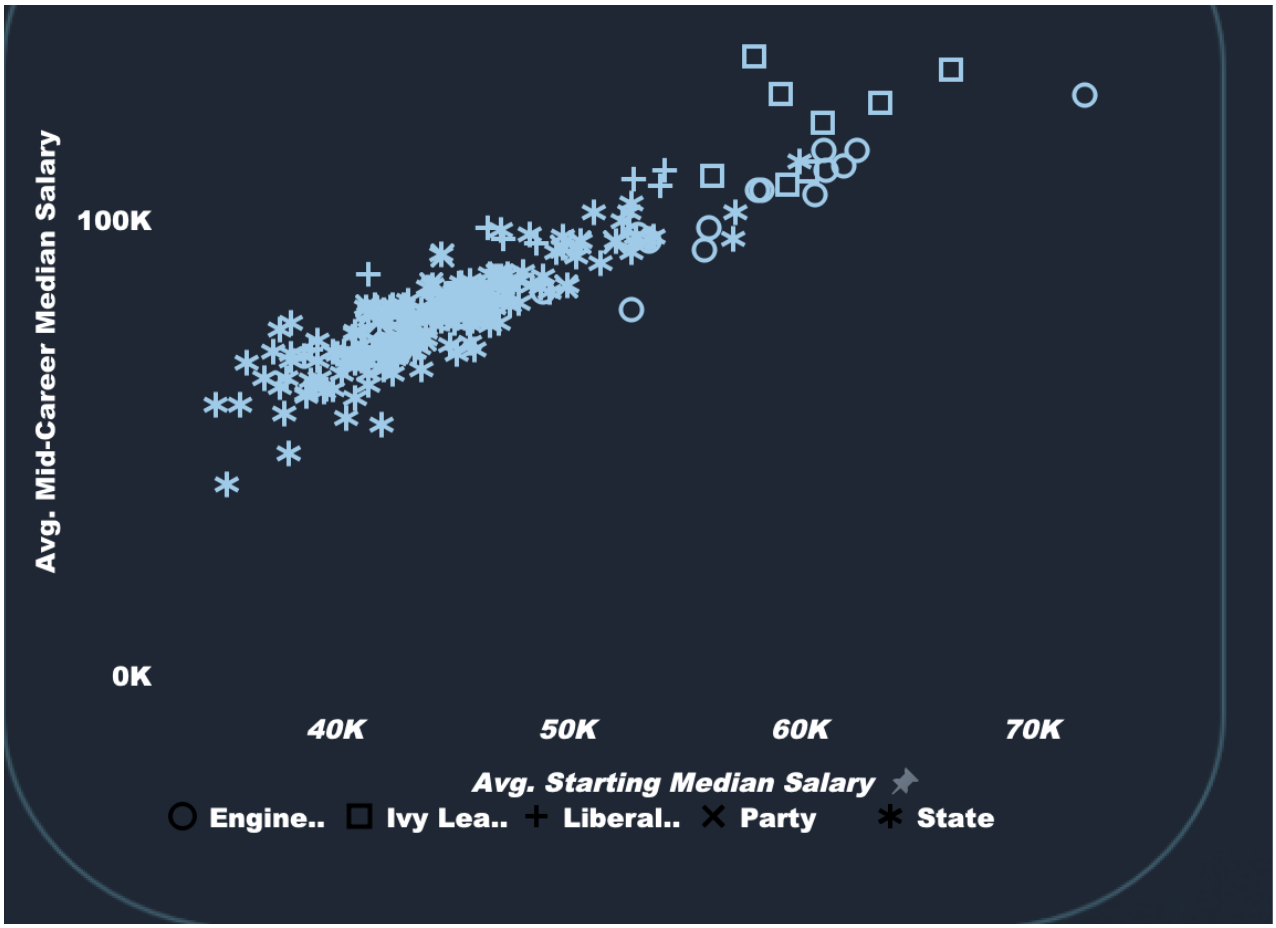
The above grouped bar plot shows the Mid-career median percentile vs school type for the 10^{th},25^{th},75^{th},90^{th} percentile. This bar plot shows that Ivy colleges are at the highest with $270k . Different shades have been used to represent the different percentiles.

School name vs starting median salaries



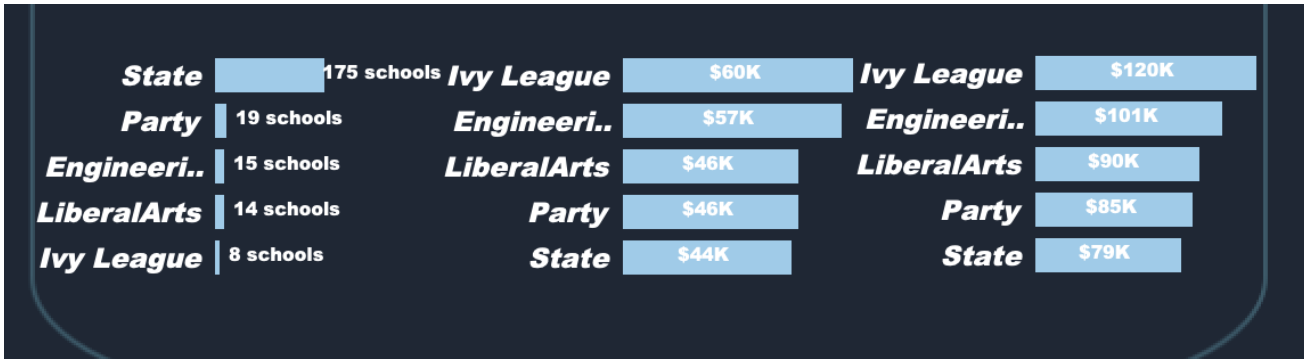
In this visualization, we plot a bar plot of the school name vs starting salaries. And also put a reference line to see the minimum to maximum salary range with minimum being $50k ,we see that the ones with starting median salary above $50k are mostly ivy league and engineering schools.

Average mid-career median salary vs Average starting median salary



The starting median salary and mid-career median are highly correlated as mid-career salary depends on how high the starting salary is along with many other factors like skills and performance. We can see the colleges with mid-career salaries above $100k . This scatter plot has different shapes for different college types , and it is clearly visible that Ivy leagues come at the top with all above $100k and state colleges come at the last.

School type vs Count of schools/Starting median salary/Mid-career median salary



The above 3 bar plots show the number of schools under each type vs 1st the count of the schools ,with state being the highest with 175 schools(76%) of the total schools then the average starting median salary and the average mid-career median salary with ivy league taking the highest in both bar plots with $60k and $120k respectively.