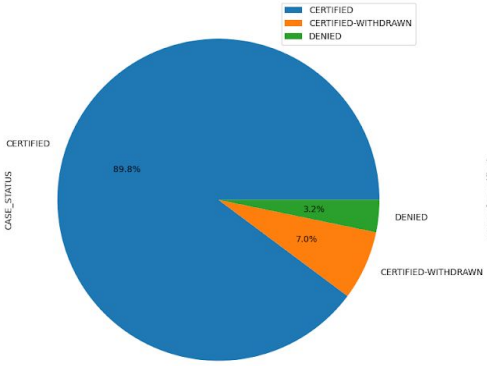
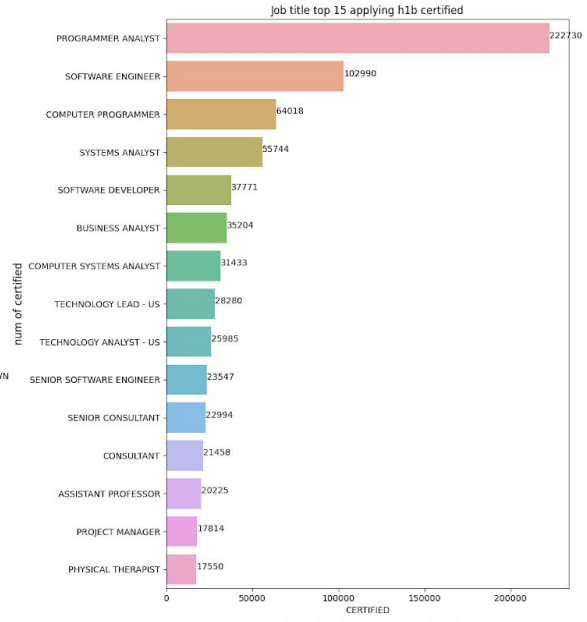
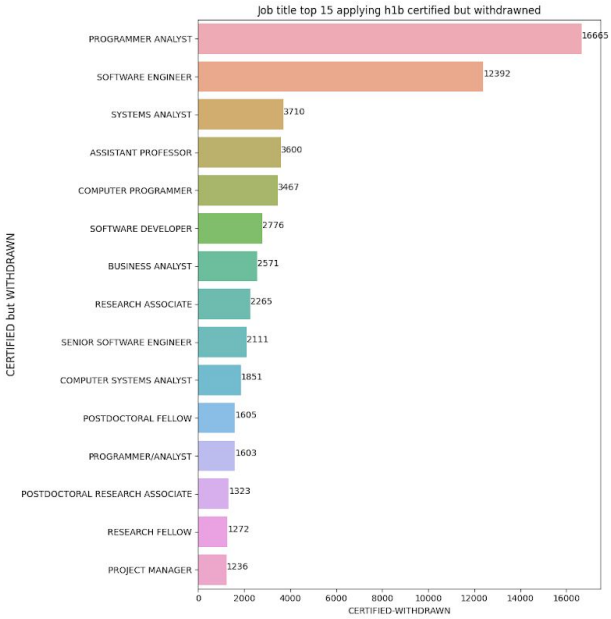
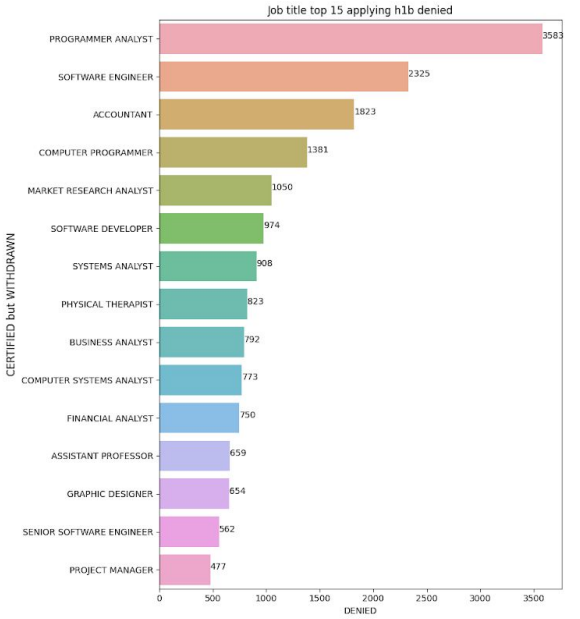
- Developer: Sophia Xiao @ 2020/1/13
- Dataset downloaded from Tianchi

  https://tianchi.aliyun.com/dataset/dataDetail?dataId=83994
- Goal:
  - Major application status (Certified/ Certified- Withdrawn/ Denied)
  - Wages of top 15 jobs of under each status
- Website for study (cited):
  - https://stackoverflow.com/questions/42770379/pandas-change-order-of-crosstab-result
  - https://stackoverflow.com/questions/38337918/plot-pie-chart-and-table-of-pandas-dataframe
  - https://stackoverflow.com/questions/29219055/plot-top-10-verse-all-other-values
- Steps:
  - Read data from csv, then create a new dataframe contains only job title and different status
  - For pie chart, count num of times each status appear
  - For bar plot, sum the num of status appear under each job title, rank them descendingly
  - Title, label and annotation
- Problems I had during the project:
  - Had little trouble putting all the plots in a figure, but them I research on website and make subplot, putting the plots in the right position (laytight())
- Saved plot









- Analysis:
    - From the pie chart, we can see most H1B applications are passed, and only about 3.2% denied.
    - The rest of three plots are mostly listed by analysis, programmers etc. Mostly coming from IT & business industries. (before I make the plot, I assume the reason of receiving deny is the choice of job, but turns out because the base

number apply H1B from these industries is huge, IT & business take all the categories)

- Despite the huge number of application might blur the result, we can still see some pattern from the plot:
  - Jobs getting passed are mostly experienced, skillful (e.g. project manager, senior consultant, etc.)
  - Accountants have a bigger denied percentage compared to certified lists.