https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics/monthly-nonimmigrant-visa-issuances.html

Problem:

Let us start with the problem we are trying to solve here.

The problem we are trying to solve here is related to the data on Visa issuance by the United States around the world. The data is made public by the U.S., but it is not exactly convenient to look it up and access it, and even for those who find it, it is not easy to skim through the entire data to find information we’re looking for.

So, the solution is to build a webapp that allows users to access this data in a drastically convenient manner.

That is the primary goal.

How did I do this?

1. Data Scraping and Cleaning
   1. pdf\_downloader
      1. Used beautiful soup to download all desired PDFs   
         (one for each month: by country and visa\_type, visa\_issued)
   2. pdf\_to\_csv
      1. used library:tabula to convert PDFs to CSVs
   3. csv\_clean
      1. Cleaned CSV
   4. data\_cleaning\_exceptions
      1. For inconsistent ……
   5. master\_df
      1. to merge all monthly data into a single master DataFrame
      2. master\_df.csv
2. Flask Webapp
3. Deployed to Heroku

What is next?

1. Frontend:
   1. Bootstrapping
   2. Basic Visualizations
      1. Trends
      2. ….
2. Multi selection feature:
   1. Select range of dates and months[slicing]
   2. Implement pivot tables

Issues/ Questions:

1. Initial Loading time [[Link](https://nivisa.herokuapp.com/)]
2. Filtering method [[Link](https://github.com/TashiNyangmi/Visa/blob/main/filter_for_app.ipynb)]
   1. Using index or column
      1. If index:
         1. both date and country [multi – index]
         2. date
         3. …other
3. Store master\_df as csv or implement some form of SQL rdbms such as SQLlite?[Tutorial [Link](https://www.youtube.com/watch?v=girsuXz0yA8)]
4. Deploy it using AWS? [[Link](https://www.guru99.com/heroku-vs-aws.html)][[Notes](#_AWS_vs_Heroku:)]
   1. If yes?
      1. For experience
      2. Better overall?
      3. …other reason
   2. If no?
      1. Reason:
5. Am I even paying attention to the right things?
   1. I know it’s not ML related. It does not even include any analytics.
   2. How useful it in terms of showcasing my skills? Initiative taking?

## AWS vs Heroku:

* Heroku is a PaaS build on Top of AWS
* Heroku is container based
* Heroku offers a ready-to-use environment
  + Deployment process of AWS is quite complicated
* Heroku for Startups, Medium Business
  + AWS for Medium and Large Enterprises
* Heroku can meet low computational demands
  + AWS can meet very high computational demands
* Heroku does not need infrastructure maintenance
  + AWS needs a dedicated DevOps guy
* Other:
  + Heroku allows easy integration w/ other AWS products
    - 1. Export Data as csv : Entire data, or filtered data, (ideally an API)
      2. How I created an API blog…..(TDS)
      3. Pivot tables not really worth…
      4. A webpage w/ my personal analysis …trends…and others…
      5. Predictions maybe? NLP…
      6. Target users: LES background, researchers….
      7. Link to Github
      8. Possible partners:
      9. People who’ve done something similar…
      10. Try Reddit…
      11. [www.immigranthistory.org](http://www.immigranthistory.org)
      12. CUNY’s the Graduate Center : Data for the Public Good fellowship program for Graduate students…