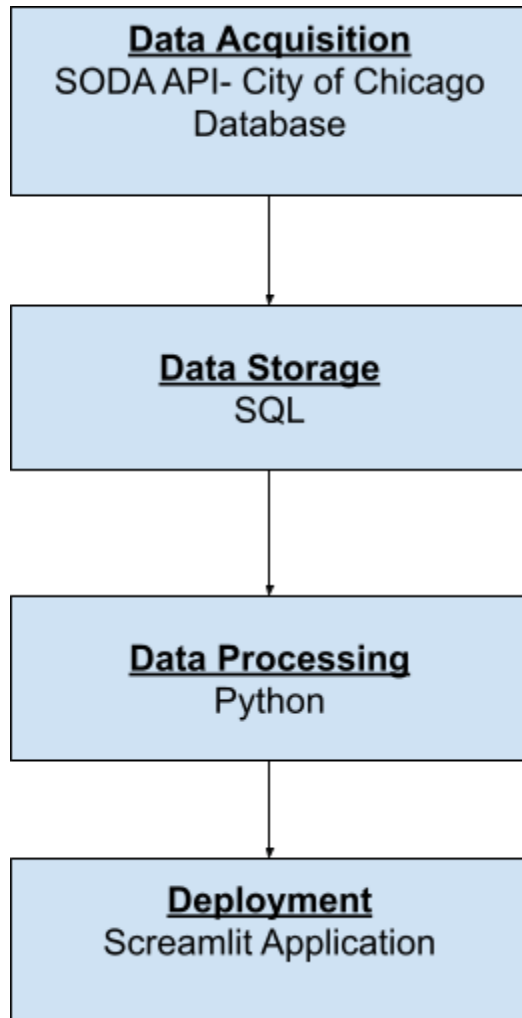Question/Need:

- What is the end service that your project will provide? What is the purpose of the system you plan to build?

I am planning to build a dashboard that looks at food inspection cases at different facilities in the city of chicago. Residents will be able to go here and look if the facility they plan to go to have a history of failing an inspection.

- Who is your client and how will that client benefits from your end service?

My client is individuals who visit various food distributing facilities in the city of chicao. This will allow for them to make a informed decision before visiting that facility.

- What does your end-to-end data pipeline look like? **You should submit a preliminary sketch of your pipeline in the form of a diagram or workflow list with your proposal.** For more directions and examples see the Data Pipeline and Project Workflow pages.

```
┌─────────────────────────────┐
│     Data Acquisition        │
│  SODA API- City of Chicago  │
│         Database            │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│      Data Storage           │
│         SQL                 │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│     Data Processing         │
│        Python               │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│       Deployment            │
│   Screamlit Application     │
└─────────────────────────────┘
```

Python
Deployment
Screamlit Application

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you're downloading, the site you're scraping, etc.)

I plan to use the City of Chicago food inspections <u>dataset.</u> I will create an api connection to the site.

- Do you plan to be able to automatically pull in new data at a regular cadence (e.g with Airflow or a cron job)?

I hope to be able to do this.

- What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?

One row of data shows:

- Inspection_id
- Dba_name
- Aka_name
- License_
- Facility_type
- Risk
- Address
- City
- State
- Zip
- Inspection_date
- Inspection_type
- Results
- Latitude
- Longitude
- Location
- violations

- What characteristics/features do you expect to work with? In other words, what are your columns of interest?

I would like to explore correlations between facility and their types of violations as well as locations of facilities with the most violations

- If modeling, what will you predict as your target?

Correlations between violations.

Tools:

- How do you intend to meet the tools requirements of the project (data storage solution and at least one tool for cloud computing, big data handling, or web deployment)?

I plan to deploy my dashboard on Streamlit and use SQL for exploratory data analysis.

- Are you planning in advance to need or use additional tools beyond those required?

Not at this time.

MVP Goal:

- What would a minimum viable product (MVP) look like for this project?

Preliminary analysis on food inspections in the CIty of Chicago over various locations and time.