

Progress Evaluation - Milestone 4

Title: Customizable Analysis and Visualization Tool for COVID Cases

Team Members:

- Calvin Burns, cburns2017@my.fit.edu (Team Lead)
- Sam Hartle, shartle2017@my.fit.edu
- Stian Olsen, shagboeolsen2017@my.fit.edu
- Nicole Wright, nwright2017@my.fit.edu

Advisor: Dr. Philip Chan, pkc@cs.fit.edu

Client: Dr. Philip Chan, pkc@cs.fit.edu

Progress Matrix for Milestone 4:

Task	Completion %	Stian	Sam	Nicole	CJ	To do
1. Continue work on scatter plot/plot utility	90%	80%	10%	0%	0%	More useful type of dataset for scatter plot
2. Continue work on plot creation tool	90%	0%	0%	0%	90%	Operations card update to be more intuitive
3. Continue work on creating lab testing plot/Start API auto update	90%	30%	60%	0%	0%	API auto update
4. Dataset management and application system	85%	0%	0%	60%	25%	User Edit/Delete data

Discussion of each accomplished task and obstacles for Milestone 4:

- Task 1:

This task revolved around finding a good use for the scatter plot and finishing up the plot utility. With help from Dr.Chan, we decided that we would use the scatter plot to see if there is a correlation between positivity rate and duration of mask mandates in Florida. We were able to calculate the positivity rate. We decided to calculate the positivity rate on a daily basis because of uneven reporting of negative cases. One challenge we encountered was that the Florida mask mandate dataset is not a useful dataset because only 10/67 counties have a mask mandate policy and the policy is still in place in these counties. One idea for scatter plot would be to do the positivity rate for cases in the US vs lockdown or international flight restrictions.

- Task 2:

This task was a continuation of work on our create plot page. The first part of the task was improving the general UI. First we added a card where the user can select the columns from the dataset. Next we added a card where the user can apply filters. Our initial version of the filter card was poorly designed and required the user to have an existing knowledge of query statements. To improve the design, we add a series of dropdowns that dynamically load data from the dataset. Dynamically loading data based on what column is selected was a very involved process. In addition the datatype of the value dropdown needed to change based on the datatype of the selected column. This too added a layer of complexity. There are three fields per filter, column name, conditional, and value. Finally, an operations card was added. This is still a work in progress and requires the user to write pseudocode that we use in our system.

- Task 3:

This task revolved around finishing work on creating the lab testing line plot that shows positivity rate over time. At this point, the operations are hardcoded and not specified in the Create Plot GUI. The operations involved in this plot (count, sum, and division) could be bundled into a positivity rate operation for the GUI in the next milestone. Obstacles included being able to format the timestamp column from the dataset into a way that would render properly on the frontend.

- Task 4:

This task created user and admin flows for applying to make a dataset public, viewing all submitted applications, and admin flows for viewing and approving submitted applications. A new database model was created, Application. This represents a user submitted application. Then a new form page was created where a logged in user can select a dataset, provide a source, and a source description. After submitting, the application is listed in the users View Applications table. The initial status is pending. An admin user can log into the admin page and navigate to the applications page. Here they can view all applications and filter based on status and date submitted. Finally, the admin can review an application and change the status. The last admin to edit an application is saved as the reviewer. Status are updated system wide. Approved datasets are made public.

Discussion of contribution of each team member to Milestone 4:

- Stian: Worked on the scatter plot and layering on the Florida map. Was not able to complete the scatter plot because of the lack of Florida mandates. Finished work on the Florida heat map. Helped Sam with the work on positivity rate vs time from lab testing data.
- CJ: Built the new Create Plot page. Updated the filters card to use 3 field format, column, conditional, and value. Helped Nicole with submit application, view applications, and the admin application pages.
- Sam: Finished work on plotting positivity rate vs. time from lab testing data. Showed the ability to chain together basic operations (count, sum, division) on the backend and process a decent-size dataset with a useful chosen statistic.
- Nicole: Built and completed the application feature. Added pages for submit application, view applications, and the admin application review process. Built the database model for Application. Wrote custom save methods for each form to save hidden userprofile values.

Task Matrix for Milestone 5:

Task	Stian	Sam	Nicole	CJ
1) Continue work on scatter plot	Find a useful dataset and plot in on a scatter plot (100%)	-	-	-
2) Operations card update to be more intuitive	-	-	-	Develop a format and UI to have operations and filters mixed together (100%)
3) Save unique workspaces	-	-	Research and develop a format for saving multiple plots in order with options (25%)	Develop a utility to render and transmit multiple plots to one page (75%)
4) Layering plots	Research and develop rules for what plots can be layered. Frontend work (75%)	-	-	Research and develop rules for what plots can be layered. Develop a utility to take two plots and combine them (25%)
5) API auto updates for datasets	-	Ensure that all datasets, both private and public, have some associated endpoint for frequent updates (80%)	-	Assist with starting point (20%)
6) Finish Application Feature	-	-	Update to use curated, shared, and private (100%)	-

Discussion of each planned task for Milestone 5:

- Task 1
 1. Mask mandate data was not readily available for FL counties
 2. Might plot volume of international flights into US vs. positivity rate
 3. Could also consider TSA daily population vs. positivity rate
 - Pull from airport specific dataset
- Task 2
 1. Move select columns to the end of the form and have their functionality applied at the end of the create plot process.
 2. Research channing operations and filters together.
 3. Rework UI to allow filters to be applied after operations.
 4. Build a better UI for operations using knowledge from research.
- Task 3
 1. Develop a format(JSON) to hold multiple plots, in order, with options(i.e. Size, colors)
 2. Create a page for users to build dashboards.
 3. Create a page to render a dashboard.
- Task 4
 1. Do research and figure out which plots can be layered.
 2. Update plot utility and javascript functions so plots can be layered.
 3. Different types of layering on Florida map
- Task 5
 1. Research API endpoints and how they will apply to our application
 2. Ensure that this updating of data occurs at a time of minimum traffic so as not to affect user experience
 3. Have system in place to determine if a dataset has not been updated from the API endpoint
- Task 6
 1. Update database models to use private, shared, and curated.
 2. Refactor codebase to use visibility instead of is_public.
 3. Update application process to work with plots and dashboards.

Date(s) of meeting(s) with Client/Advisor (same) during Milestone 4:

- January 15th (Project Plan)
- February 1st
- February 15th

Client feedback on the current milestone: See Faculty Advisor Feedback below

Faculty Advisor feedback for Milestone 4:

- Notes from 1/15/2021
 1. Look into SQL query syntax and potentially see if it could be a good way to organize operations
 - SELECT colNames or some function on a column
 - FROM tables
 - WHERE filter (e.g. limiting rows)
 - GROUP BY
 - ORDER BY
 2. At least make the male female example work using GUI and specified filters/operations
 3. For lab testing plot positivity rate vs. date
 - Issue with plotting timestamp column
 - Django does not see timestamp as a Datetime
- Notes from 2/1/2021
 1. Mask mandate data on scatter plot (Stian)
 - Plotting mandate duration vs. positivity rate
 2. Map display (Stian)
 - Coloring (heat map)
 3. For line graph, plot positivity rate vs. time (Sam)
 - Ratio operation on backend
 - Hardcode first before GUI
 4. UI improvements (CJ)
 - Use English instead of operators
 - Display range for numerical values as they type
 - Find min/max values
 - Error checking
 - Store metadata for tables to determine type
 - Pandas has method called dtype
 - Returns list of datatypes
 - Dictionary passed to frontend
 5. Admin review tool (Nicole)
 - Mock-up GUI
 - Research application submission process
 - Backend supporting tool
- Notes from 2/15/2021
 1. Scatter plot idea
 - Volume of international flights into US vs. positivity rate
 - TSA daily population vs. positivity rate
 - Airport specific data

2. Map display
 - Positivity rate, positives per capita
3. Find more frequently-updated CSV dataset for lab testing
4. UI
 - “Filters” before “Select Columns”
 - Select Columns (for output; change wording) should be last step
5. Number of days where positivity rate > 5
 - CREATE (and populate) positivity_view(date, positivity_rate)
 - #6
 - Alternatively add column (Pandas way)
 - SELECT date, count(positive) / count(id)
 - #5
 - FROM lab_testing_table
 - #1
 - WHERE positive == yes
 - #2
 - GROUPBY date
 - #3
 - ORDERBY date
 - #4
 - This will create output with two columns
 - Date and positivity_rate
6. Continuation of example from 5
 - SELECT count(date)
 - #3
 - FROM positivity_view
 - #1
 - WHERE positivity_rate > 5
 - #2
7. SQL example above illustrates two issues to consider
 - How to group operations
 - Ordering execution of operations
8. Curated vs. shared datasets
 - Curated has been approved by admin
 - Shared is any dataset a user can share with anyone else
 - At your own risk