

# CrimeTracker

Team Falcons : Phillippe, Piyush, Mack

## Sprint 1

Sep 30, 2022

created github repo  
We want to make it public

invited both to github as collaborators

Talk about assignment

2 things to start off

think what we want the tech stack to be

what do we want the data to be

get data, import it into

datasette, a tool for exploring and publishing data  
we looked at the demo

we will use datasette for data presentation

SQLite

We need reasoning for each part of the tool we need

for the toy application, use datasette to show the data

create a hard-coded example

later on, we want the whole application with mvp, user can look at any piece of data and visualize whatever

do we want to create our own dataset, or use something from kaggle  
looking at a few possibilities for data

Crimes in US communities

excellent dataset

why we think this is interesting

Seeing hotspots for crime, even visualizing them is interesting to see,  
look into live data source for crime data, this might just be a few cities

someone confirm with prof to see if this dataset is good

tech stack

lots of the backend stuff is taken care of datasette wise

We need to get the datasource, put it in a database, use datasette to create a  
visualization

The bulk of the work is creating the surrounding UI for the visualization, UI should  
have a way to upload the data,

input into sqlite database, call to the db and send to datasette

frontend where the user can upload the data, any data they want, also having a  
selector for what visualization they want

<https://www.kaggle.com/datasets/michaelbryantds/crimedata>

containerization  
using docker is fine, just create a dockerfile

prelim project development plan  
how we want to split up the work

how that looks like for each milestone

action items  
look more into datasette  
see what we can do, how easy it is to interface, how to cast the data, what format

think about more features the app should have  
maybe a selector for different visualizations

for UI, we can use react, see what we can do

backend is probably bulk of the work, use python probably

<https://datasette.io/for/websites>  
<https://datasette.io/for/data-journalism>  
<https://datasette.io/tools/download-tiles>

## Oct 06, 2022

Oct 6

Discussed what we want to make, reiterated last meeting's meeting notes.

What we need to do,

Describe tech stack, toolchain, product development plan, toy application, setup docker

Piyush created a google doc that we can write our A1 information in.  
[https://docs.google.com/document/d/1u1XdaumMEYbFADvhUH5Uu6EIC-aksnFrHDb\\_AWDGVYE/edit](https://docs.google.com/document/d/1u1XdaumMEYbFADvhUH5Uu6EIC-aksnFrHDb_AWDGVYE/edit)

Discussed toy application. We will have very basic functionality.  
We just want to load the data into a sqlite db

Wrote the product development plan

Discussed the three milestones we want to have, wrote information about it in the google doc

Piyush to work on the toy application saturday and sunday  
Philippe to work on google doc

Piyush + Mack to work on toy application and unit tests and docker

## Oct 09, 2022

==October 9 Meeting==

Time: 12:00

Participants: Mack, Piyush

==Ideas for Meeting==

The test will be an SQL query.

Piyush has used SQL in general, Postgresql

==Plan for this long weekend==

Piyush will finish the SQL database, and make a test editor. []

Mack will attempt to make a Docker container out of it. I will try to make two separate containers.

==Other notes==

Write documentation for the docker run and docker build.

# Sprint 2

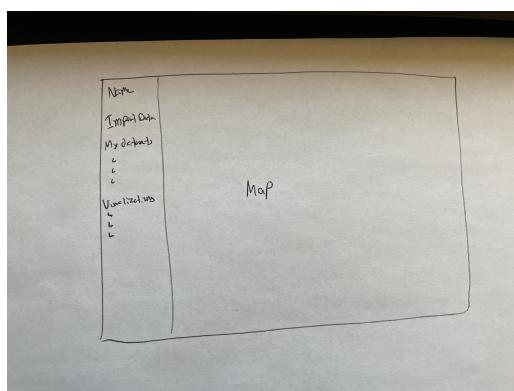
Oct 18, 2022 18:00 (Blameless Postmortem)

- Discussed Monday's lecture
- Go over the A2 assignment description
  - We want to have our Docker working properly
- What goals have been achieved?
  - Simple dummy application to build on
  - A rough plan for our milestones
  - Tech stack and reasoning behind each technology planned out
- What goals have been missed, and by how much?
  - The containerization of the application was not met. Instead, we automated using a shell script as a temporary solution
  - Splitting up the work for the milestones among ourselves
  - Project planning board setup. For example, a Kanban board in our GitHub repository.
- How plans will need to be adjusted, based on remaining time available and project scope.
  - The plan is feasible. Additional features were given low priority to ensure we complete the essential features.
- What, if any, adjustments or decisions need to be made in light of what we have learned about project scope and capacity since?

Steps

=====

1. Uploading Dataset through React Dashboard which gets stored in SQLite Database
2. Show a quick preview
  - a. Show the schema
  - b. Show a couple of rows similar to df.head
  - c. Show some stats about columns similar to np.describe
3. Visualizations: The user can select what type of visualization (map, etc.) and then click generate which sends a call to the Datasette backend
  - a. Option to check or uncheck certain columns from appearing in visualization
4. Formatting Data (Modifying SQLite Database): The user can choose to delete certain columns or rows based on the value



Oct 22, 2022 13:00

Milestone 2 Recap

=====

Mainly focused on creating a frontend with MVP functionality and providing a basic mechanism to connect to the backend.

Backend in the current state is quite minimal. Each front end task assigned will be responsible for back end tasks.

For task #4, we can create the functionality, however, the won't be a UI until we have task 2 completed. Reuse part of the quick preview ui to show the entire database. #4 to be assigned in the future.

Backend Status

=====

Nonfunctional.

Roles

====

Piyush to work on React Dashboard UI and the upload button to store the dataset in an SQLite database.

Philippe to work on the visualizations. Philippe to decide which additional visualizations to add

Mack to work on the frontend and deciding points to connect to the backend. Backend drafts such as the schema will be made.

4) Should column deletion based on nulling. For UI, we can accommodate nulls by not showing them (or using a counter)

Issues

=====

Due to midterms, we may need to delay working until Wednesday. I will handle the current case, such as Figma.

**Oct 22, 2022 22:00**

Review on our current code

**Oct 31, 2022 18:00**

Proposal:

We decided to visit at 22:00 on a daily basis for 15 minutes.

**Nov 01, 2022 22:20**

Piyush:

Docker done

Mack:

Tested Docker on Linux

Made considerations of Github Actions after looking at documentation.

Python unit tests (no PyTest)

Set up Docker container.

M: Python linter

Philippe:

Need to study the ui maps.

Proposal:

RE diagram to organize our dataset and to identify which parts require particular analysis. We need to consider how crime data will work and the sources of the crime data. Pick that specific dataset rather than blindly choose which dataset to use. How can this dataset be generalized?

**Nov 02, 2022 22:00**

Piyush:

Mack:

Philippe:

Readme:

Discussed that docker image is good information. We will eliminate the exact package count. Mack has added a trace of a successful Docker run on Linux.

Piyush Tomorrow:

Flask and available data

How proper pull request should work

Mack Tomorrow:

Minimal RE diagram + Skeleton Yaml File

**Nov 03. 2022 22:20**

**Piyush:** Finished the frontend. Needs to do the backend, uploading the files.

**Philippe:** Got the map working, except for those in Washington

**Mack:** Get GitHub actions working

## Blameless Postmortem

TBA, see Oct 18

**Nov 4, 2022 22:00**

We have combined the branch, related to the static directory and the flask server.

### Issues:

- Map data not set correctly (404 error)
- Fixed that on Docker (Philippe)
- Made changes to the Dockerfile (Philippe)
- Need to install the CURS library using pip install

### Review process:

- It has been decided that only one person is required to approve for the pull request to be effective.

### Mack:

- Approved the changes from the mulliganaceous-patch branch.
- Mack will have to prepare for a major second-round interview from Intel

### TODO:

- MyDatasets Tab and Github Actions (Mack)
- Schema (Mack)
- A2 Writeup (Submit by Sunday, finalize by Monday)

**Nov 7, 2022 13:00**

Mack has completed his Intel interview, on a reinstalled Ubuntu machine. He has not encountered any difficulties on his technical part of the interview.

**Nov 12, 2022 22:15**

Mack has discussed with the team members in regard to the extension. The extension can go up to Monday, in which grading will start. Mack has had trouble with Node.js recently, and has trouble fixing it even with the assistance of his intern manager (he is currently also working on an internship assignment). The source has to do with dependency conflicts, broken packages, and not being run sudo node. He will continue to fix the loose ends shortly before the hard deadline.

Nov 13, 2022 22:04

**Participants:** *Mack, Philippe, Piyush* (22:10)

I have pinned the Docs for the notes and for the A2. Mack will be the principal notetaker for the remaining meetings. He has read through *Software Engineering: Theory and Practice* by Shari Lawrence Pfleeger. After reading this textbook, he has a clearer idea on proposing which technologies are to use.

Tonight, we pointed out the last remaining parts of our writeup, provided documentation on how to run the application, both on Docker and locally, and listed the nine major risks of development.

Mack will work on the remaining goals for A2 that was missing, while the remaining two will work on the README.md for local and Docker development.

## Sprint 2'

The purpose of this short one-week session is to recoup our loss for Sprint 2. So far, I saw a 14 out of 20, so we got a **70%**. We need to discuss with Michael Hoye to prove we have recouped our losses and are now back on track. The reason is that there are major difficulties. Furthermore, our image is too big simply from running SQLite.

### Issues

The staff pointed out four aspects of failures which led to lost progress.

1. Image is too bulky for just running sql-lite, please consider revisiting the packages listed as dependencies for the container
  - a. Identify NPM dependencies
  - b. ETC: 2 days
2. Installation eventually succeeds, albeit with a lot of warnings.
  - a. We need to make considerations of the warnings.
  - b. ETC: 2 days
3. Backend status is non-functional. We need a functional backend at this stage of assignment.
  - a. This is something we missed. We need to research more on backend.
  - b. ETC: 5 days
4. Test fails since it is not able to query the backend database.

Sun Nov 20th, 2022 19:45 ~ 20:45

At the time of installation, Docker is no longer using. We are using MUI, React Router, and React Native. We do not know what is causing the anomalous behavior, so we have to investigate.

#### **Sprint III Meeting Times:**

- Extended discussions are conducted for new ideas, overall progress and journaling, and trying out our program.
  - It should last at least an hour and there should be an agenda.
  - Saturday and Sunday evenings and nights are the best place for extended discussion. They are done online.
- Scrums are conducted every business day. It is a report on what updates and difficulties one have encountered.
  - After running scrums on 22:00, we later moved to 21:00.
  - We will now move the scrum earlier to **19:00**. They are done online.
- In-person meetings are done occasionally and are conducted based on common in-person events and critical issues.
  - A possible time is in-person this Thursday 7pm.

### This week's schedule:

- **Falcon**
  - TBA
- **Piyush**
  - Can work on this project any time this week that isn't 10am to 4pm Thursday and 1pm to 2pm Friday.
- **Mack**
  - Will dedicate full time on this project from Wednesday to recoup
  - Taking a lighter course load as well

### Urgent

Due to Docker issues, we need to discuss more regarding the Docker installation. We need to clarify Docker installation. How can you prove the backend is functional? It serves frontend, and responds to frontend API.

I tested locally, and it is not working. I am not able to get the backend working after installing the frontend. *Falcon is a Mac, Piyush is a Windows, and Mac is a Linux!*

### Projected Goal

We need to show beforehand that our Docker is working, and identify installation instructions and subtleties we missed upon installation. We have to identify why people are complaining about the troubles they are seeing.

Deadline: Nov 22

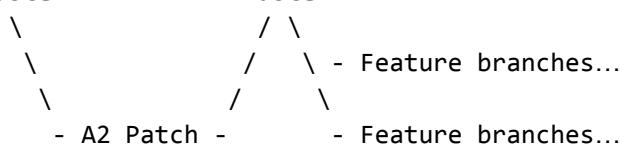
### Testing and Automation

Due to previous issues, it is hard for Mack to continue testing our app. We will carry on testing and automation throughout A3, in which logging will also be combined with testing.

### Workflow

We will propose that our development branches will update.

Master ----- Master ----- A3



## Blameless Postmortem (Assignment 3)

What goals have been achieved?

- React Dashboard setup
- Flask backend setup
- Import Dataset functionality
- Map visualization functionality
- Docker containerization setup

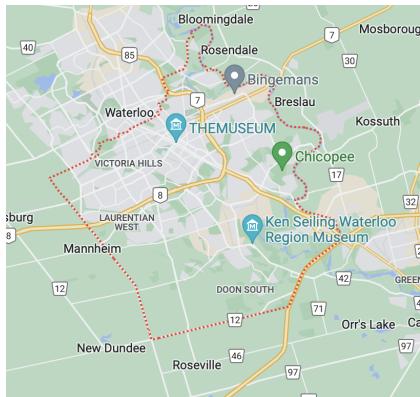
What goals have been missed, and by how much?

- My Datasets tab: Viewing list of currently imported datasets and their schemas
- Testing: Need to do this on both frontend (Jest) and backend (Python Unit tests)
- Automated Testing: GitHub Actions integration

How plans were adjusted, in light of that information, including goals or features that were dropped, what other changes were made and why?

- My Datasets tab, testing, and GitHub Actions will be carried over from A2 to A3 due to teammate's external commitments.

## A3 Work: Look into accuracy of reverse geolocation



### Piyush

- Improving Dashboard UI
- Adding more options for Importing Datasets as mentioned in A2 Writeup
- Formatting Dataset (CRUD operations)
- Visualization Selector: Adding options for Visualizations (Collaborate with Philippe)

### Philippe

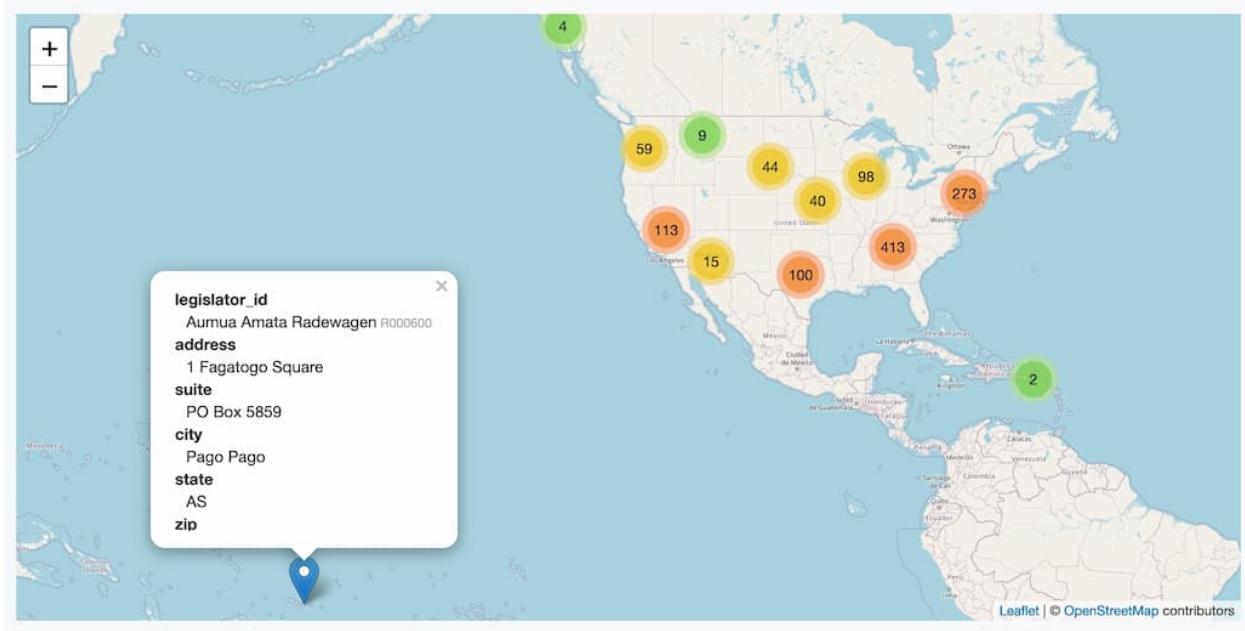
- Reverse geolocation fix
- Map boundaries
- Improve data ingestion workflow
  - When dataset is submitted, export the server response immediately and cache it
  - If modifications are made to a dataset, update the cached version
  - This will create a much more responsive website
- Live data (stretch goal)
  - Find a way to scrape data from <https://twitter.com/TPSOoperations>
  - Take intersections and add those as points on the map
  - Live data can be used to update current backend

### Mack

- Logging, Testing, Automation
- My Datasets tab: Getting Datasets list and Displaying Schemas
- Sorting data by crime type
  - Data can be color coded to identify crime hotspots
  - Crime hotspots helps us identify where the crime is frequent in and causes of crime.
  - Total crime
- Filtering crime type
  - Felonies vs Misdemeanors
  - Allow user to filter the dataset by columns and values (essentially not have the user write SQL, but rather let them filter using a React form)

## Undecided

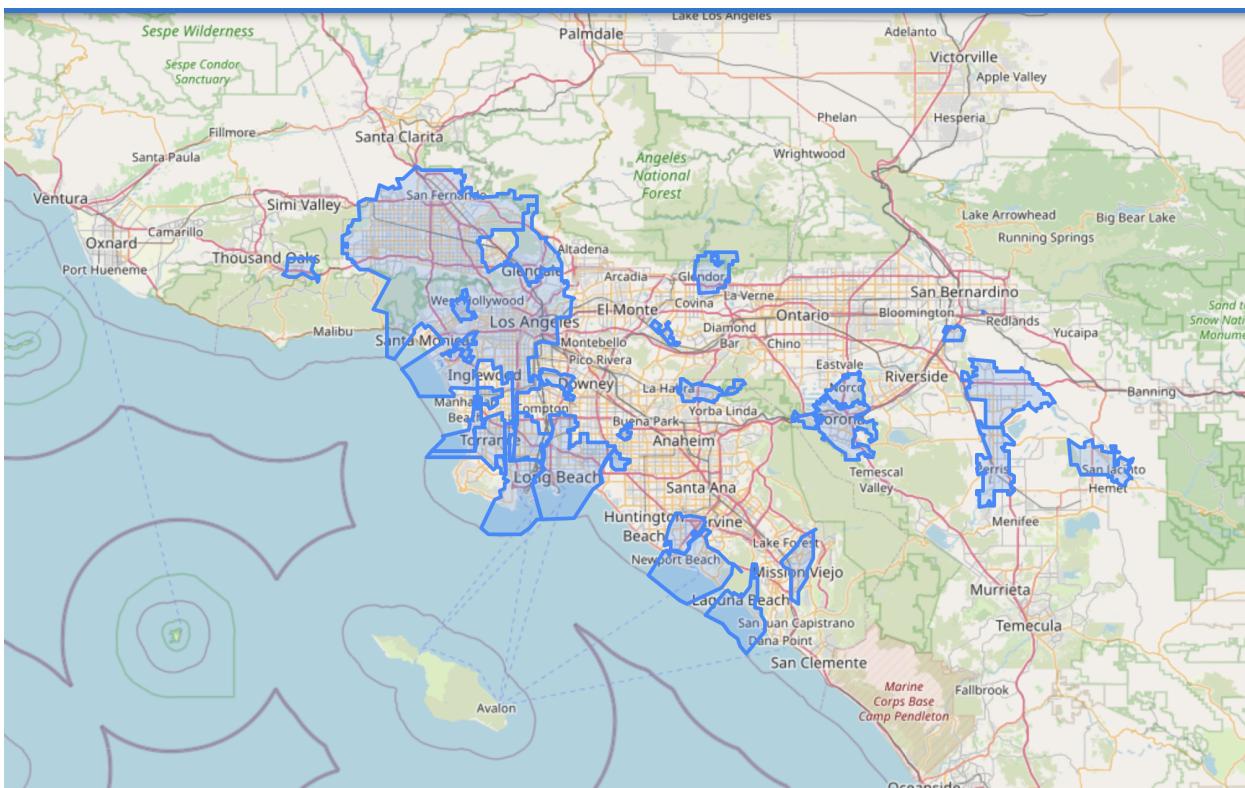
- TBA



## Monday Nov 28 meeting

Some progress has been made.

Philippe has been able to add map boundaries for our locations on the map



Data takes a really long time to process, the api we are using, openstreetmap Nominatim is rate limited to 1 request per second. Our dataset is very large, so processing all that data will take several hours.

As this takes a long time, we will need to cache our results. Provide a sample .db or .csv that can be used to create our data.

Data ingestion method will be as follows: Receive csv -> transform to sqlite db -> make appropriate requests from Open Street Map, modify sqlite database to have that information -> Serve requests by outputting db in json format.

Data from that api is not always correctly formed, for example lat and long are sometimes swapped, city names do not exist, etc.

Was able to detect swapped lat and long, but for city names that don't exist, we will have to leave that out of our dataset.

Unit tests should ensure the validity of our datasets

Philippe to use json.parse to take in geogjson data in the frontend

## Nov 29 meeting

Mack is experiencing issues with communicating with the backend from the frontend.

Issue is due to port differences. Mack was running the backend on port 5000 when the frontend makes a request to port 3000.

Discussed changes to the mapData api endpoint. Mack will need to add dbname endpoint to his requests to mapData.

Github actions still in progress

## Dec 01st, 2022 11:50

The three contributor's have merged their branches and all issues have been merged.  
Issue: Nonunique identifier. Each square on the map is its own div.

!MULLIGANACEOUS!!: to import, or multiple datasets, it generaes one db file for each. for each, os grabs all db files

[02:10]!!MULLIGANACEOUS!!: aggregate; do it from backend

[02:12]!!MULLIGANACEOUS!!: state data: backend or locally frontend

[02:13]!!MULLIGANACEOUS!!: i tink that can be done fron backend

[02:16]!!MULLIGANACEOUS!!: letting user select data

[02:17]!!MULLIGANACEOUS!!: 24 man hours to get the heatmap and

```
[02:17]!!MULLIGANACEOUS!!: you cannot do  
[02:18]!!MULLIGANACEOUS!!: testing and logging  
[02:18]!!MULLIGANACEOUS!!: testing on the react and standard unit test for backend
```

## Dec 2

Mack and Philippe conducted code review on Mack's pull request.  
A rudimentary my-datasets tab was merged into master.  
Mack's request to change the logo was rejected  
Github partially functional github actions were merged into master.  
The docker test is passing, python unit tests are not functional at this time as Mack has not written them.  
The design of how to detect new databases was discussed. Mack will create a new endpoint that runs ls in the database folder and returns the result.

## Dec 3

Agenda:

- Merging in finished My Datasets tab and Visualization Selector
- Ensuring testing suite is complete and all tests pass
- Ensuring Docker and GitHub actions pipeline are running successfully
- Creating a template of our presentation and assigning slides

Discuss with Mack about his failing db requests.

Mack did not specify datasetID

Created template for slides.

Decided the following slides will be made

- Dataset/Problem we chose
- Solution we chose to develop (visualization dashboard)
- Tech stack
- Sprints overview
- Demo
- Technical Challenges, Solutions, and Considerations
- Future improvements

Demo will be screenshare, ~1 minute

Philippe will be responsible for demo, technical challenges and solutions regarding the map, sprints overview

Mack will be responsible for technical challenges, solutions, and considerations, future improvements

Piyush will be responsible for Dataset/Problem we chose, solution we chose to develop, tech stack.

Full runthrough will be scheduled for Dec 4 at 9PM.

## **December 7**

As the instructor changed the presentation deadline to Dec 8, the full runthrough scheduled for Dec 4 was moved.

Status update on unit tests

Mack was not able to develop any unit tests

Philippe discussed offline with Mack and took the python unit test task over. Philippe was able to write some unit tests ensuring the validity of the data ingestion portion. The unit tests are passing and there is a green checkmark on github.

Mack's latest pull request

There were several bugs introduced in Mack's latest pull request.

- The map view would load several datasets at the same time erroneously.
- The map view showed state and city data at the same time preventing city data from interactions.
- The heatmap thresholds were modified causing areas of high crime to be marked as less crime.
- The violent crime calculations were modified to include non-violent crime.
- A new heatmap color of gray was added.

We discussed the issues that needed to be fixed. Mack will revert the heatmap

changes, fix other bugs. Mack said the pull request will be merged tomorrow

Mack will work on his slides as well.

A rehearsal is scheduled for tomorrow (december 8) at noon

## **December 8**

Agenda:

- Merge Mack's pull request
- Rehearse for the presentation

Mack was not present for the scheduled meeting

Philippe and Piyush reviewed Mack's pull request. The heatmap changes were not reverted. The file structure was changed causing unit tests to fail. Philippe and Piyush worked to fix these issues and were able to merge the pull request.

As Mack was not responding to messages, a rehearsal was done. Piyush agreed to take over Mack's part of the presentation if he did not attend.

Some suggestions were made to improve the slides, the slide order was reorganized.

Another meeting will be scheduled later today if Mack becomes available.

## **December 8, meeting 2**

Agenda:

- Rehearse the presentation
- Last minute changes if required

Mack responded to messages and was available for a rehearsal.

Multiple rehearsals were conducted, some suggestions were made to improve Mack's part.

Slides were finalized.