## **ETL Project**

|  |  |
| --- | --- |
| Project Title: | California Wildfires & Bay Area Air Quality |

## **Team Members**

|  |
| --- |
| **Names:** |
| Nabil |
| Grace |
|  |
|  |

## **Project Week Timeline**

|  |  |
| --- | --- |
| Date | Subject |
| S 12/1 | Project/ ETL Intro; project, its goals, and requirements, identify feasible projects, submit a project proposal. |
| T 12/4 | By end of day: have loaded both data sources into a database |
| TH 12/5 | By end of day: have merged data sources successfully. |

## **Project Description**

We will be looking data on California wildfires, and comparing it with data on Bay Area air quality.

|  |
| --- |
| **Finding Data**  Your project must use 2 or more sources of data. We recommend the following sites to use as sources of data:   * [data.world](https://data.world/) * [Kaggle](https://www.kaggle.com/) * [Data.gov](https://www.data.gov/) * [Public API Listing](https://github.com/toddmotto/public-apis)   You can also use APIs or data scraped from the web. However, get approval from your instructor first. Again, there is only a week to complete this! **Data Cleanup & Analysis** Once you have identified your datasets, perform ETL on the data. Make sure to plan and document the following:   * The sources of data that you will extract from. * The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc). * The type of final production database to load the data into (relational or non-relational). * The final tables or collections that will be used in the production database.   You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process. **Project Description** Your Project Description |
| The kind of data you'd like to work with/field you're interested in (e.g., geodata, weather data, etc.)  **Response**: Data on California wildfires, data on Bay Area air quality  The kinds of questions you'll be asking of that data  **Response**: Are California wildfires a major influence on Bay Area air quality? How significant an influence are they? How much of our air quality can be explained simply by fires? Are wildfires getting worse, and is our air quality getting worse at all?  Possible source for such data  **Response**: We’ve identified datasets on California wildfires and also on Bay Area air quality. |

**Project Report**

## **Project Report - Due Saturday 10/20**

At the end of the week, your team will submit a Final Report that describes the following:

* Extract: your original data sources and how the data was formatted (CSV, JSON, MySQL, etc).
* Transform: what data cleaning or transformation was required.
* Load: the final database, tables/collections, and why this was chosen.

## **Data Sources**

|  |  |  |
| --- | --- | --- |
| # | Description | URL or Resource Link |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |

## **Finding Data APIs to be Used**

|  |  |  |
| --- | --- | --- |
| # | API Description | API URL |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |

## **Repo Link**

|  |
| --- |
| Repo URL |
|  |

## **Retrospective Notes (10/20):**

**Team Reflection**:

* What Went Well
* What Did not go well
* What can **WE** improve for next time

**Team Member Feedback (directed at your team member(s))**:

* What Went Well
* What Did not go well
* What can **YOU** improve for next time

|  |
| --- |
| **Team Reflection & Team Member Feedback** Google Doc Link |
|  |

**Self Reflection**:

* What Went Well
* What Did not go well
* What can **I** improve for next time

|  |  |  |
| --- | --- | --- |
| **Self Reflection** Link |  |  |
| Name |  |  |
| Name |  |  |
| Name |  |  |
| Name |  |  |

## 

## **Stuck time ( How long until you consult a teammate?):**

* 30 min?
* 1 hour?

## **Working Agreements:**

We will both attend class on Tuesday and Thursday.

Nabil will project manage and let Grace know what needs to be done.

Each of us will upload one datasafurce to a database before Tuesday.

Sunday evening from 8pm to 8:15pm we’ll talk on the phone to see how we are. We’ll decide on next steps based on stats at that time.

**EXAMPLE:**

During Sprint Do The Following:

* Team members attend daily stand ups @ 6:30pm (T,TH) 10am (sat)
* We will be expected to be working on Sun and M,W,F (is this true?)
* Should a team member have a conflict, s/he/they updates the slack team channel in advance of the meeting
* We believe in in the value of planning collectively as a team
* Every Team Member is engaged and involved.
* Every Team Member is committed to the value of the application over individual recognition.
* Every Team Member will practice active listening
* Every Team Member will be directly engaged with the work - not answer texts or phone calls, social media, or other off-topic material (with the exception of emergencies -which will be communicated to the team)

|  |
| --- |
| **Your Team Working Agreements** |
| During Sprint Do The Following:   * We will both attend class on Tuesday and Thursday. * Nabil will project manage and let Grace know what needs to be done. * Each of us will upload one datasafurce to a database before Tuesday. * Sunday evening from 8pm to 8:15pm we’ll talk on the phone to see how we are. We’ll decide on next steps based on stats at that time. |

**Notes:**