**DELIVERABLES + TIMELINES**

1. **Initial Draft of Findings** - due by beginning of Lesson 11
   * Describe the users represented in the data. Do you think it’s representative of the population of Firefox users?
   * Describe any initial findings related to your research questions;
   * Describe what you still need to do for your presentation in order to make a sound recommendation;
   * Format: A four-five paragraph e-mail to your boss.

1. **10-minute Presentation** - given to class during Lesson 12
   * Presentation outlining your findings and your recommendation;
   * This will be a final presentation to your boss and team at work that will build on your earlier e-mail and summarize your overall findings and recommendation;
   * Format: Google Slides or PDF (Keynote/PPT need to be exported)
2. **SQL Queries + Questions** - due by beginning of Lesson 12
   * Include the questions you’ll be looking to answer with the data;
   * For each question, include the corresponding query or queries that were used to answer them, as well as the output of each query
   * Format: .sql file

*\*A detailed description of each deliverable can be found below in the “Requirements” section*

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**PROMPT**

You step into a Doctor Who Police booth and travel back to the year 2010.

Promptly you encounter a stressed out junior product manager working at Mozilla Firefox.

They tell you they have been tasked to decide if the new version of Firefox should have an improved bookmark experience or should they find a way to support larger number of tabs.

Considering the analytics workflow, you decide to look at a [recent test flight](http://datahub.io/dataset/a-week-in-the-life-of-a-browser-version-2/resource/ec631c92-c975-4549-b646-74bb2ec2d778?inner_span=True) with customer surveys and usage analytics to better inform your thinking before deciding which of the two areas to recommend.

*\*Note: The data required for this project can be queried from the AWS server you have been using in class, but please read through the “recent test flight” link above for the official documentation.*

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**REQUIREMENTS**

**1. Initial Draft of Findings:**

* Describe the users represented in the data.
  + Do you think it’s representative of the population of Firefox users?
* Describe any initial findings related to your research questions.
* Describe what you still need to do for your presentation in order to make a sound recommendation.

**2. 10-minute Presentation:**

* Summarize the data in your presentation
  + Present your findings, results, and recommendations to the class
  + Review the Users table. Summarize the users represented in the survey.
  + Review the Survey table. How many of the total users completed the survey?
  + Of users that completed the survey, identify the number of users who are new to firefox and who are long time users.
  + Include other information that is important to include in your initial findings that may have affected your recommendation and final presentation.
* Summarize your exploration of the bookmark usage feature in your presentation. Some sample questions you could answer include:
  + What’s the median number of bookmarks? What’s the average number?
  + What fraction of users launched at least one bookmark during the sample week?
  + What fraction of users created new bookmarks?
  + What's the distribution of how often bookmarks are used?
  + How does number of bookmarks correlate with how long the user has been using firefox?
* Summarize your exploration of usage of browser tabs in presentation. Some sample questions you could answer include:
  + What's the distribution of maximum number of tabs?
  + Are there users who regularly have more than 10 tabs open?
  + What fraction of user have ever had more than 5 tabs open? What fraction of users have ever had more than 10 tabs open? What fraction of users have had more than 15 tabs open?
* Compare the two features and create a coherent recommendation to your boss.
  + Do you agree or disagree with your opinionated colleague?
  + Do you think the team should go in a different direction?
* Include analysis of data with statistics and outliers
* Include recommendation with predicted outcomes and identified action items
* Present limitations, assumptions
* Identify follow-up problems and questions for future analysis

**3. SQL Queries + Questions:**

* List the appropriate business questions you were trying to answer
* Use correct, efficient syntax for the functions and operators needed
* Include corresponding output listed in the .sql file

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SUGGESTED WAYS TO GET STARTED**

Look to the [Analytics Workflow Document](https://drive.google.com/open?id=0BwdiPd53UHiSWjVxeUZLRUliUWc) to guide you through each step:

* Identify and Understand the Problem - Considering the prompt, think generally about the problem and how you might want to approach it. Ideally, what type of data would you have to make a decision and produce a recommendation?
* Access and Obtain the Data - Your company, Mozilla, is running their own PostgreSQL server, and the engineers that designed the survey have added that data to three tables: events, survey, and users. Connect to the appropriate server using the software of your choice (pgAdmin III or DBVisualizer, for example).
  + To connect to the server, please use the following login credential with the firefox\_database:
    - url/host:
      * West-coast (includes Austin and Chicago) - analyticsga.cuwj8wuu6wbh.us-west-2.rds.amazonaws.com
      * East-coast - [analyticsga-east2.c20gkj5cvu3l.us-east-1.rds.amazonaws.com](http://analyticsga-east2.c20gkj5cvu3l.us-east-1.rds.amazonaws.com/)
      * APAC and Europe - [analyticsga-apac2.csuojbfcexhv.ap-southeast-2.rds.amazonaws.com](http://analyticsga-apac2.csuojbfcexhv.ap-southeast-2.rds.amazonaws.com/)
    - port: 5432
    - username: analytics\_student
    - password: analyticsga
* Understand the Data **-** Your colleagues in the engineering department are very organized and have written documentation for the product managers who need to use their databases. Make sure you read it! [Link to documentation](https://web.archive.org/web/20160304073326/https://testpilot.mozillalabs.com/testcases/a-week-life-2/aggregated-data.html)--*note that the live site has been taken down but the web archive works great; there are also copies of these pages and datasets in the Project 2 folder*. HINT: make sure to identify which user actions will be of most use to your research questions. Browse the tables and ensure you know how the data is organized and how the tables relate to each other. Answer basic questions about the data. When presenting the results of survey data, it’s often good to inform your audience about the characteristics of the surveyed users. Do you think these users are representative of all Firefox users? Why or why not? What other limitations might exist with this data?
* Prepare, Structure, and Clean the Data - In order to be able to determine your recommendation and create your presentation, you will need to write a number of queries to prepare, structure, filter, clean, and aggregate your data.
* Analyze and Visualize the Data - In order to determine findings and create a recommendation, you will want to use statistical (mean, median, mode, min, max, etc.) functions as well as other aggregate functions (count) to summarize your data. Export your data to Excel to create any visualizations that will help you either explore your data or provide evidence for your recommendation.
* Present the Results of Analysis - Based on your findings, create a coherent argument about which project is more worthwhile: tabs or bookmarks.
* Create a PowerPoint presentation that summarizes your findings and makes a recommendation.
* Make sure to include results of your analysis and any visualizations that help boost your argument.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**RESOURCES**

* Specific Class Decks on how to use SQL and access the database
* Dataset [documentation](https://web.archive.org/web/20160304073326/https://testpilot.mozillalabs.com/testcases/a-week-life-2/aggregated-data.html) *(also found in Project 2 folder)*
* Recent [Test Flight](https://datahub.io/dataset/a-week-in-the-life-of-a-browser-version-2/resource/ec631c92-c975-4549-b646-74bb2ec2d778?inner_span=True) data
* Data Mining [Framework](https://decisionstats.files.wordpress.com/2011/10/12345.png)
* [Analytics Workflow Document](https://drive.google.com/file/d/0BwdiPd53UHiSWjVxeUZLRUliUWc/view)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**EVALUATION**

* Your presentation will be graded using the requirements section above as a rubric.
* The Excel worksheet containing the cleaned data will be evaluated using the requirements section above as a rubric.