CIS 9440 - Data Warehousing and Analytics

Class #11

Class note #1

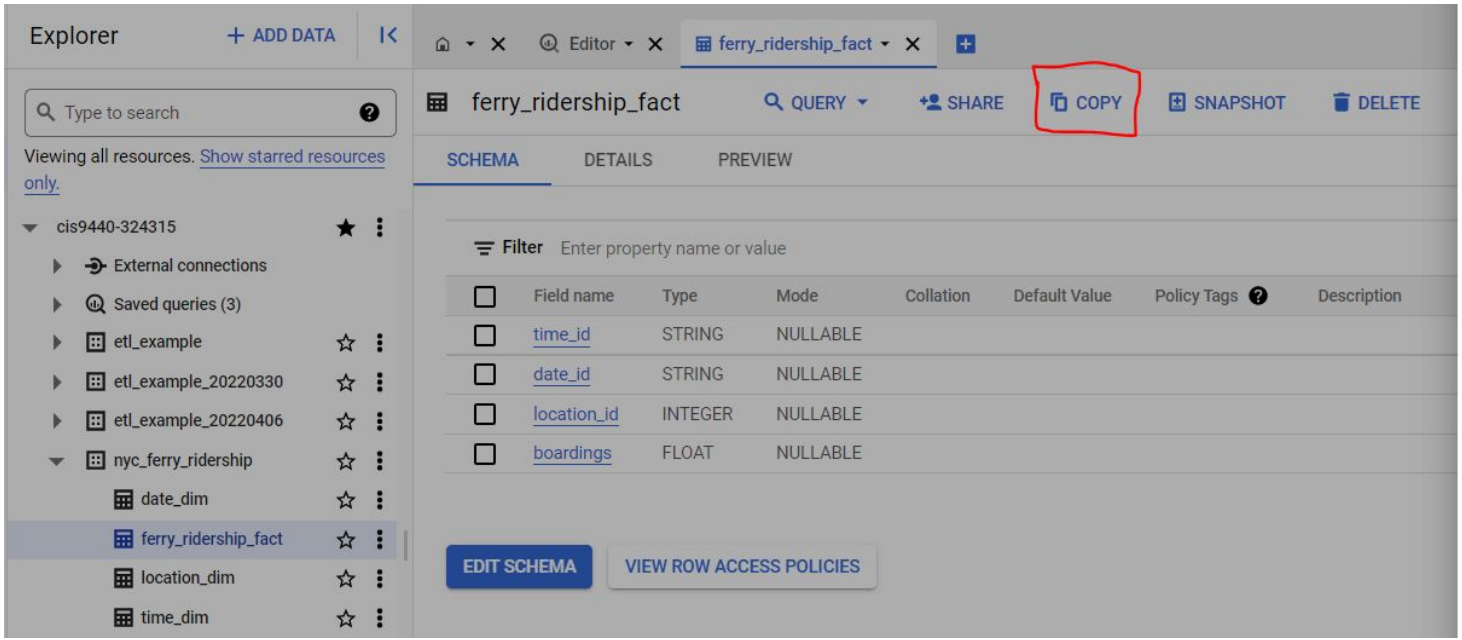
Thanksgiving Eve class next week, 11/23/22, will be hosted remote via Zoom. No in-person class on this day.

Class note #2

How to share datasets on BigQuery?

1. Grant your teammates as permissions access (with the Share button) to be both "data viewer" and "data editor" (use your teammates gmail addresses when doing this).
2. Go to each table you want to share
3. Copy the URL when you are in the "details" view of the table
4. Send the URL to your teammates
5. Your teammates click on the link, then ensure on the top-right of their screen that they are logged into the correct google account (the same one you granted them access with, sometimes when you click on the link it takes you to a different google account)
6. Your teammates may then query/view the table(s)

Class note #2



Source

Project name
cis9440-324315

Dataset
nyc_ferry_ridership

Destination

Project *
cis9440-324315

Dataset *
nyc_ferry_ridership

Table *
ferry_ridership_fact_copy

Unicode letters, marks, numbers, connectors, das
the specified destination table if needed.

Advanced options

Week 11 Class Overview:

1. Tableau warmup
2. Types of BI Applications
3. Different BI Users
4. BI Workshop #2
5. What is Git?

Week 11 Class Overview:

1. **Tableau warmup**
2. Types of BI Applications
3. Different BI Users
4. BI Workshop #2
5. What is Git?

Download the data

All Taylor Swift songs:

https://docs.google.com/spreadsheets/d/1UpllEgGTxPMqeEeWm0q_-KxhQ4Dai91T/edit?usp=sharing&ouid=100275929541227111171&rtpof=true&sd=true

Tableau Warm Up together

1. **Contribution Analysis**: Total Minutes by Album
    a. First, as a Pie Chart
    b. Then, as a Tree Map
        i. Add Label information
        ii. Remove Color

2. **Time-Series Analysis**: Total Songs by Year

Tableau Warm Up individually

1. **Comparative Analysis**:

   Avg Track Mins by Track Number

2. **Distribution Analysis**: Songs by Track Seconds

Week 11 Class Overview:

1. Tableau warmup
2. **Types of BI Applications**
3. Different BI Users
4. BI Workshop #2
5. What is Git?

BI Applications, what are the differences?

A BI Application is a BI Platform deliverable that is developed to for specific users to measure a specific KPI.

Examples:

- Dashboard
- Scorecard
- Report
- Data Visualization
- Ad hoc query
- Predictive Model

When to use a Report?

- broad, historic, and **static** data.
  - Reports are not very flexible.
- recurring basis: daily, weekly, monthly, or quarterly.
- wide audiences - like an entire division of an organization
  - Reports may be drilled in on different data by business unit
- data in a table format

When to use a Dashboard?

- **multiple KPI's** in a singular view.
- **data that will be interacted with**, drilled upon, or manipulated by the users.
- near-live data, and used on a more frequent basis than a Report.
- unlike reports, Dashboards do not quickly deliver a message. They allow the user to find the most important messages.

When to use a Scorecard?

- measuring performance against a benchmark.
  - Example: Current Sales vs Quarterly Sales Target
- more specific than reports, with static views of very specific KPI’s
- Scorecards may be for wide or narrow audiences.
- Like reports, Scorecards are not very flexible.

When to use a Data Visualization?

- generated ad hoc, for specific presentations, emails, or research.
- One place you may find a Data Visualization is a Team presentation to the company.
    - targeted purpose and message
    - Oftentimes, a Data Visualization is created from a Dashboard. In this case, it’s a specific cut from a Dashboards full data.

Poll!

Ideal BI Applications

Week 11 Class Overview:

1. Tableau warmup
2. Types of BI Applications
3. **Different BI Users**
4. BI Workshop #2
5. What is Git?

Who are the different BI Users in your organization?

Business Intelligence (BI) is leveraged by many different people at an organization. From salespeople, to data scientists, to the marketing team, the BI Platform is used to make data-driven decisions.

It is the BI Designers job to **ensure the BI Applications are usable by their intended audiences**!

> Example: Marketing KPI's as Reports, not Analytical Sandboxes

Different BI Users

- Casual Consumers
  - Account Executives, Marketing team, Store Teams, HR associates, Graphic Designers, etc.
- Analysts
  - Financial Analyst, Sales Analyst, Marketing Analyst, Operations Analyst
- Power Users
  - Data Analyst, Business Analyst, Business Information Team, Business Insights Team, Marketing Analytics Team
- Data Scientists/Data Science Team

BI Users: Casual Consumers

- Looking for static reports in the exact format they need
- Can navigate simple dashboards
  - May require dedicated training session
    - BI Training always encouraged!
    - Best equipped to use BI Applications with an **onboarding** training session
- Will ask Power Users for new BI Applications

BI Users: Analysts

- Analysts can extract great insights from many BI Applications. They are inquisitive, and **relay great findings to their greater teams**. They enjoy BI functionality like:
  - Filtering
  - Drilling
  - Manipulating
- Analysts typically do not grasp the complexity of the Data Model within the Data Warehouse. Thus, they are not BI Application builders, but they are users.

BI Users: Power Users

- Power Users are best equipped to work with the BI Platform
  - They often have **BI Platform certifications**
  - They usually have special BI access to “write” or “edit” Applications and in the BI Platform
- Power Users are allocated to **develop and maintain the organization's BI Applications**
- **Warning**: Without Power Users delivering all BI Applications the organization needs, the organization will create “data shadow systems” that serve as short-term solutions to subsets of an organization, but long-term problems to the entire organization
  - Chapter 16 in your Textbook is about Shadow Systems

BI Users: Data Scientists

- Each organization may have a few Data Scientists
  - They may have varying titles
  - Or, there may be a Data Science team
- Data Scientists often use/build more advanced BI Applications like predictive models or analytical sandboxes
- Predictive Models as BI Applications require special setup, and are often ad hoc projects

Poll!

Casual Consumers

10 Minute Break

Week 11 Class Overview:

1. Tableau warmup
2. Types of BI Applications
3. Different BI Users
4. **BI Workshop #2**
5. What is Git?

BI Workshop #2 - situation

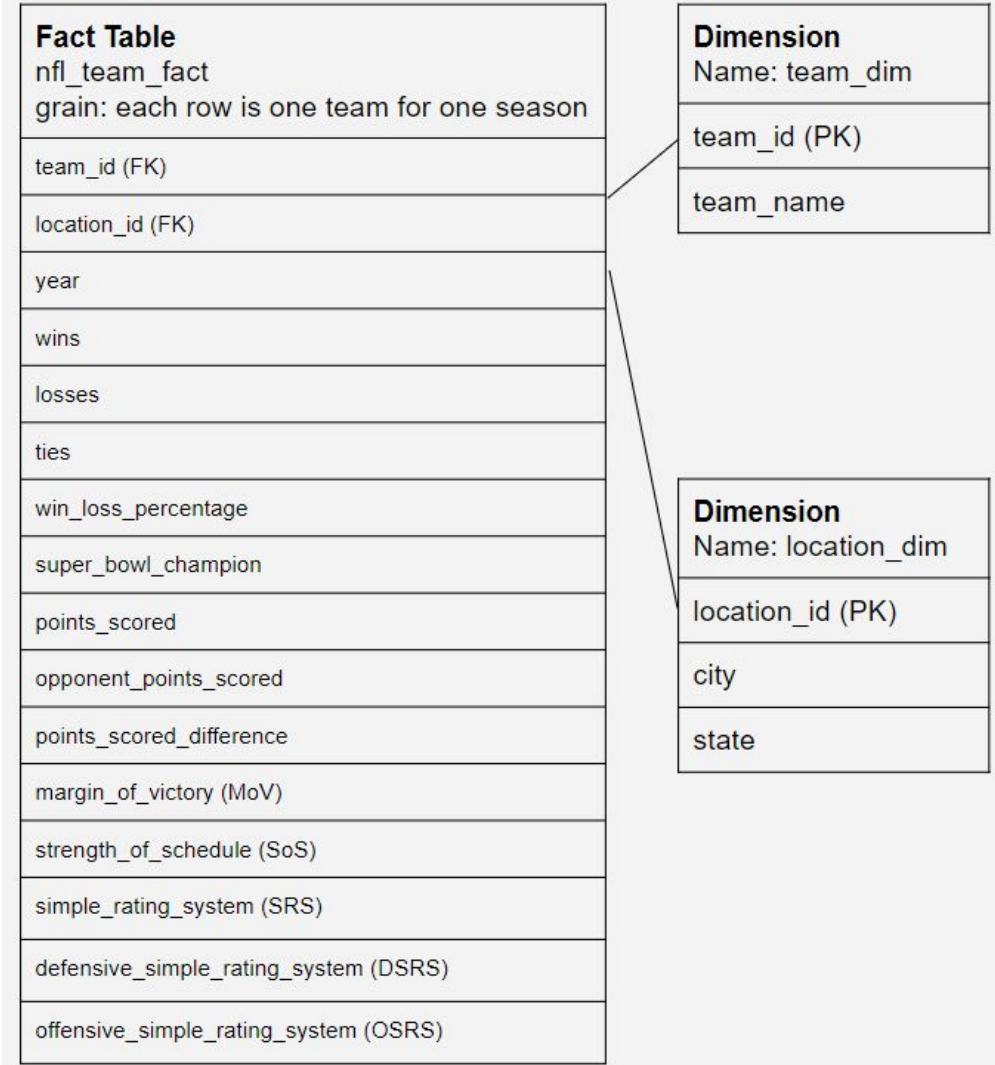
- Situation - A group of investors is choosing an NFL team to invest in, they hire you to build a Dashboard to aid their decision
  - **Audience**: Executive Group of Investors
  - **Deliverable**: dashboard
  - **Type of Analytics**: prescriptive

BI Workshop #2 - project guidelines

- Situation - A group of investors is choosing an NFL team to invest in, they hire you to build a Dashboard to aid their decision
  - Create 3 KPI's
  - Create 1 Tableau Visualization (Worksheet) for each KPI
  - For each visualization, explain why you used that type of visualization
    - Example: used a histogram to show a distribution analysis of total wins by team location. Added a tooltip because...
  - Display your 3 Tableau Worksheets on 1 Tableau Dashboard
  - (Optional) - publish your Dashboard to Tableau Public
  - Come to a conclusive recommendation and submit this Google Form

BI Workshop #2 - Dim Model

- Download as an excel file here
- Data starts in Super Bowl era, starts in 1967



Week 11 Class Overview:

1. Tableau warmup
2. Types of BI Applications
3. Different BI Users
4. BI Workshop #2
5. **What is Git?**

GitHub/portfolio

From our python/SQL/coding work together and your own work from other courses, you may choose to post your work on GitHub. This way, you can link your GitHub on your resume (if needed) and use specific pieces of code as portfolio examples.

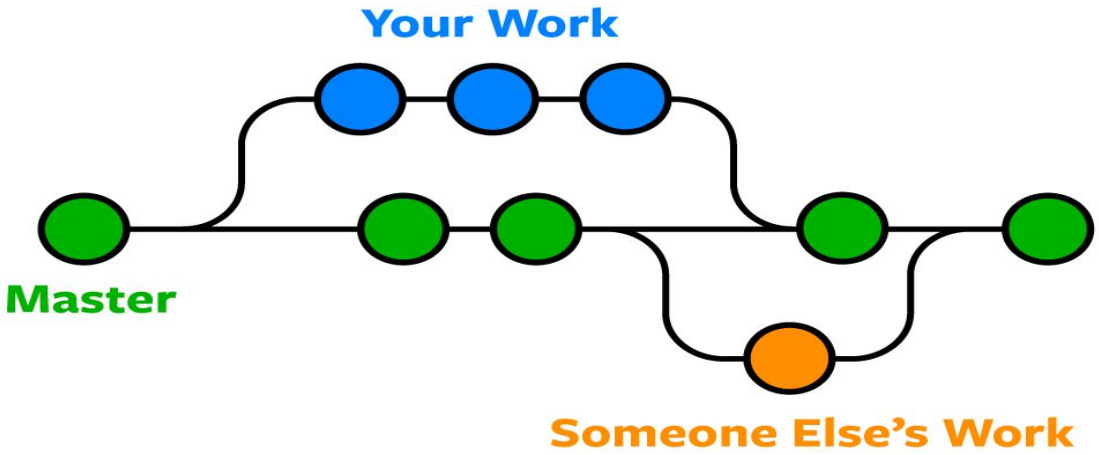
You may also apply for a Student Developer GitHub, which has many great features:

https://docs.github.com/en/education/explore-the-benefits-of-teaching-and-learning-with-github-education/use-github-for-your-schoolwork/apply-for-a-student-developer-pack

GitHub example: https://github.com/odonnell31/SQL_interview_practice_generator

What is Git?

Git is software for tracking changes in any set of files, usually used for coordinating work among programmers collaboratively developing source code during software development. Its goals include speed, data integrity, and support for distributed, non-linear workflows.



Using Git

- Initialize Repositories
- Create Branches
- Commit Changes
- Merge Branches to Master

Git Cheat Sheet

Git: configurations

```
$ git config --global user.name "FirstName LastName"
$ git config --global user.email "your-email@email-provider.com"
$ git config --global color.ui true
$ git config --list
```

Git: starting a repository

```
$ git init
$ git status
```

Git: staging files

```
$ git add <file-name>
$ git add <file-name> <another-file-name> <yet-another-file-name>
$ git add .
$ git add --all
$ git add -A
$ git rm --cached <file-name>
$ git reset <file-name>
```

Git: committing to a repository

```
$ git commit -m "Add three files"
$ git reset --soft HEAD^
$ git commit --amend -m <enter your message>
```

Git: pulling and pushing from and to repositories

```
$ git remote add origin <link>
$ git push -u origin master
$ git clone <clone>
$ git pull
```

What is GitHub?

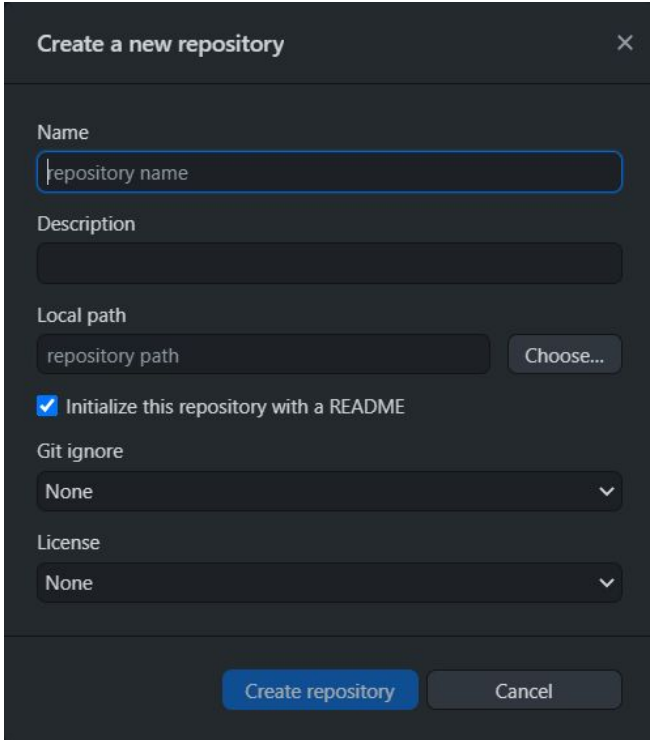
GitHub is a provider of Internet hosting for software development and version control using Git. It offers the distributed version control and source code management functionality of Git, plus its own features.

GitHub Alternatives

- BitBucket
- GitLab
- AWS CodeCommit

Create a GitHub Account

1. Create a free GitHub account: link
2. Apply for a “Student Developer Pack”
    a. Go to https://education.github.com/
    b. Click on “GitHub Student Developer Pack”
3. Download GitHub desktop:
    a. https://desktop.github.com/
4. Create a new Repository ---------->
5. Go to the filepath of your new repository
    a. Create a folder named “img”
        i. Put all project images here
6. In GitHub desktop, click “Publish Repository”



Homework:

1. Final Project Milestone #4 on Blackboard, due Friday, 11/25/22
2. Final Project Milestone #5 on Blackboard, due Friday, 12/7/22