Colab / SQL Reflection

Group: Jason Feng, Katie Williamson, Meet Patel, Tianyi He

Summary

Colab

- Google Colab is a collaboration platform that allows developers to work together on a document that includes chunks of text and code
- For text cells, Colab uses LaTeX, and thus follows the same formatting syntax as Markdown in other technologies
- Colab integrates with Google Drive to allow users to easily share with their collaborators and keep revision history. However, unlike Google Docs, two people cannot edit the Colab document simultaneously
- Colab also allows users to add comments to code and text boxes, just like they can in Google Docs, so that teammates can add questions or suggestions to a colleague's work without actually changing it

SQL

- SQL (Structured Query Language) is the standard language for relational database management systems (RDBMS) and is useful in handling organized data composed of variables and relations between different variables of the data
- SQL allows us to interact with relational databases through queries. These queries can allow us to perform a number of actions such as insert, update and delete information in a database
- In class, we used SQL to analyze grocery sales by doing the following:
    - Setting up the environment
    - Exploring the tables and relationships
    - Building the SQL query
- We answered the following types of questions in lab using SQL queries:
    - Identify the aisle has the most products
    - Identify the aisle has the most products ordered
    - The most frequently ordered products with names that contain "Diapers"

Key Implications and Takeaways

Colab

- Colab is a simple, easy to use tool that provides a convenient option for us to collaborate on projects that involve both coding and accompanying business context / analysis
- The integration of text and code blocks allow users to clearly explain their work, which may make it a more effective tool for collaboration with business users that have less technical / coding expertise
- However, since the tool doesn't allow multiple users to edit a document at the same time, users have to communicate clearly about who owns editing at any given moment

SQL

- SQL is a simple yet powerful language that is used widely as a business intelligence tool
- To answer key business questions, users can follow the following steps:
    - First, identify relevant relational tables and common entities across tables
    - Then, write a query to join different tables and isolate relevant entities
    - Finally, modify the query to calculate the relevant metric and result

Additional Resources

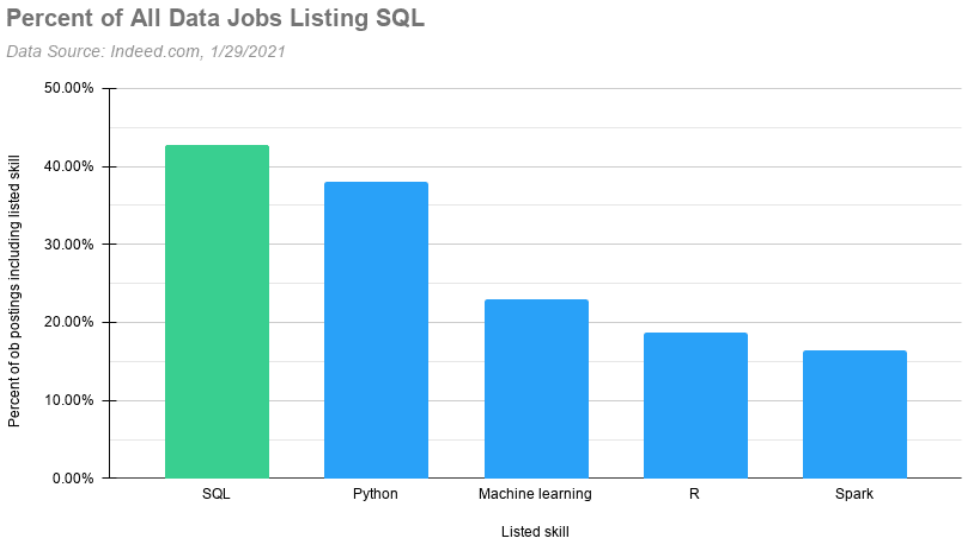
Colab

- There are many sample Colab notebooks that demonstrate machine learning applications such as the machine learning examples below. These examples can be studied to better understand how to use Colab for machine learning
    - Demo for using Nvidia's NeMo conversational AI Toolkit to swap a voice in an audio fragment with a computer generated one[1]
    - Classify IMDB movie reviews as either positive or negative[2]
- There are many tips to use Colab more efficiently. The article below provides a few recommendations including top keyboard shortcuts[3]
- To further understand how to use data in Colab including loading data and visualization, you can refer to the link below for detailed descriptions[4]

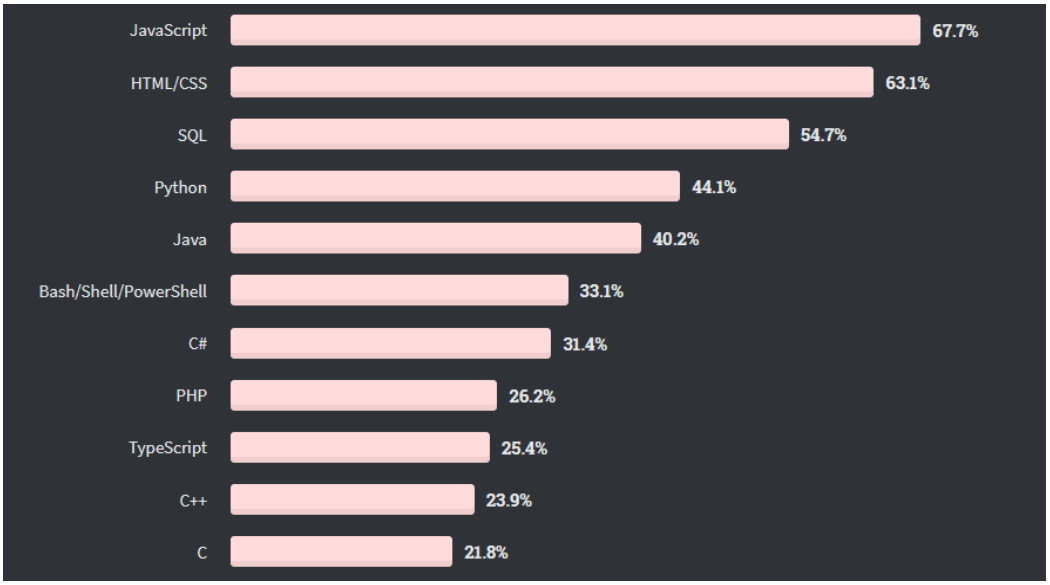
SQL

[1] NeMo voice swap app - Colaboratory (google.com)
[2] basic-text-classification.ipynb - Colaboratory (google.com)
[3] 5 Amazing Google Colab Hacks You Should Try Today (analyticsvidhya.com)
[4] Welcome To Colaboratory - Colaboratory (google.com)

- SQL has been a highly sought after skill for a variety of jobs ranging from data analysts, business analysts, engineers, database administrators, to data scientists. In fact, as this chart shows, of more than 32,000 data jobs on Indeed, SQL was the most in-demand skill of all data postings[5]



  - Additionally, SQL is one of the most used languages in all of the tech industry



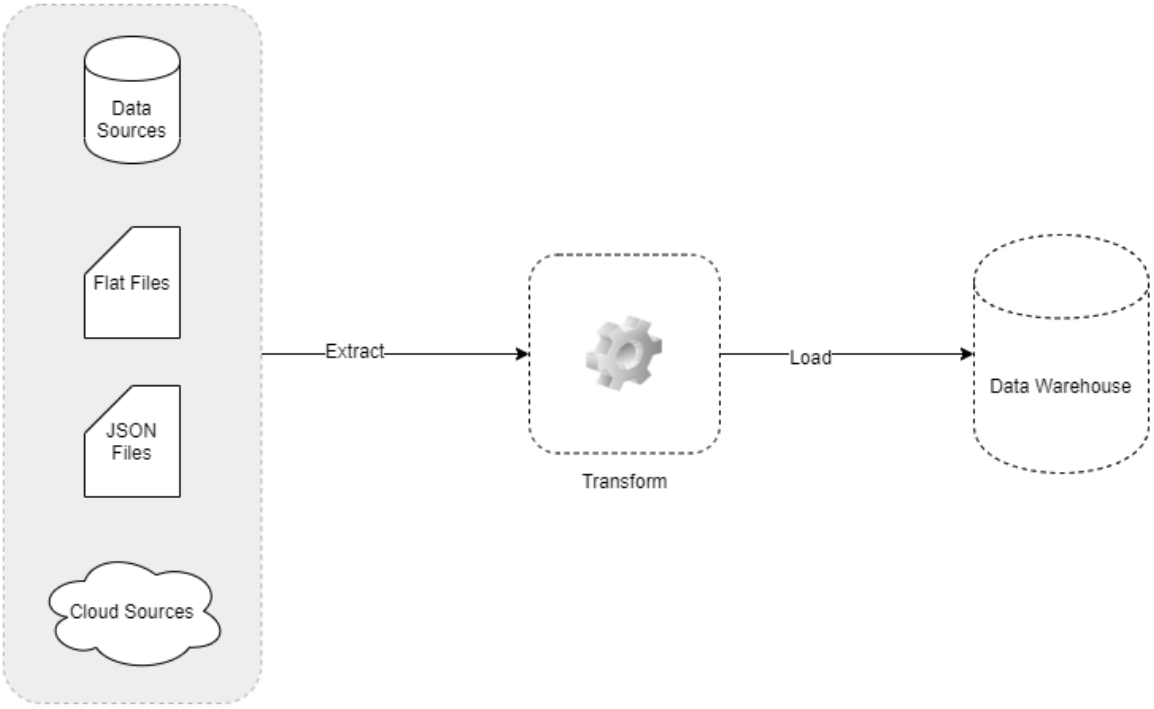
- There are, however, several different versions of SQL. Each version with different syntax and functions. This table illustrates some differences:[6]

[5] Why You Need to Learn SQL If You Want a Job in Data (2021 Update!) (dataquest.io)
[6] SQL Server, PostgreSQL, MySQL... what's the difference? Where do I start? - DataCamp

	SQL Server	MySQL	PostgreSQL	SQLite
SELECT ...	Select [col1], [col2]	SELECT col1, col2	SELECT col1, col2	SELECT col1, col2
Data from tables is case sensitive?	Yes WHERE name = 'John' Or WHERE name = 'john' are not the same	No WHERE name = 'John' Or WHERE name = 'john' are the same	Yes WHERE name = 'John' Or WHERE name = 'john' are not the same	Yes WHERE name = 'John' Or WHERE name = 'john' are not the same
Using quotation marks	name = 'John' only	name = 'John' or name = "John"	name = 'John' only	name = 'John' or name = "John"
Aliases for columns and tables	SELECT AVG(col1)=avg1	SELECT AVG(col1) AS avg1	SELECT AVG(col1) AS avg1	SELECT AVG(col1) AS avg1
Working with dates	GETDATE() DATEPART()	CURDATE() CURTIME() EXTRACT()	CURRENT_DATE() CURRENT_TIME() EXTRACT()	DATE('now') strftime()
Window functions i.e., OVER(), PARTITION BY()	Yes	Yes	Yes	No (need to use subqueries instead)

    - There are these different SQL dialects because, although there is a SQL standard, different database management systems have added their own functionalities[7]
- A major use for SQL is through the ETL process: Extract, Transformation, and Load
    - This technique allows SQL to connect to source data, extract that data from those sources, transform the data in-memory, and finally load the altered data into the data warehouse[8]

[7] How To Find Your Way Through the Different Types of SQL | by Marie Lefevre | Towards Data Science
[8] An overview of ETL and ELT architecture (sqlshack.com)



- Other specific commands, such as those listed in the table below, allow SQL users to communicate effectively with the database to perform different functions

Command	Description
CREATE	Creates a new table, a view of a table, or another object in the database.
ALTER	Modifies an existing database object, such as a table.
DROP	Deletes an entire table, a view of a table or other objects in the database.
SELECT	Retrieves certain records from one or more tables.
INSERT	Creates a record.
UPDATE	Modifies a record.
DELETE	Deletes a record.
GRANT	Gives a privilege to users.
REVOKE	Takes back privileges granted from users.

[9]

- SQL joins allow for the merging of different data sources based on common columns

[9] What Is SQL & How Does It Work? A Guide to Structured Query Language | Springboard Blog