SQL & Python for Analytics

USC Annenberg Digital Lounge
Week 5: Analytics & Visualization with SQL and Python

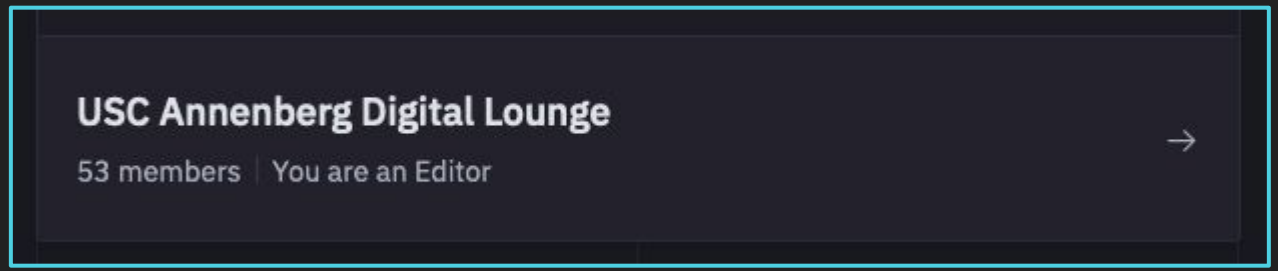
Today’s session

This is going to be interactive, so please be ready to try out the exercises as we go!

We are going to be using a tool called Hex, which is a tool for using SQL and Python for doing data analysis and visualization.

Shout-out to Hex for giving us their Professional Plan for free!

Go to app.hex.tech and sign up with your USC email address or Microsoft login. When asked to select a workspace, choose Annenberg Digital Lounge:



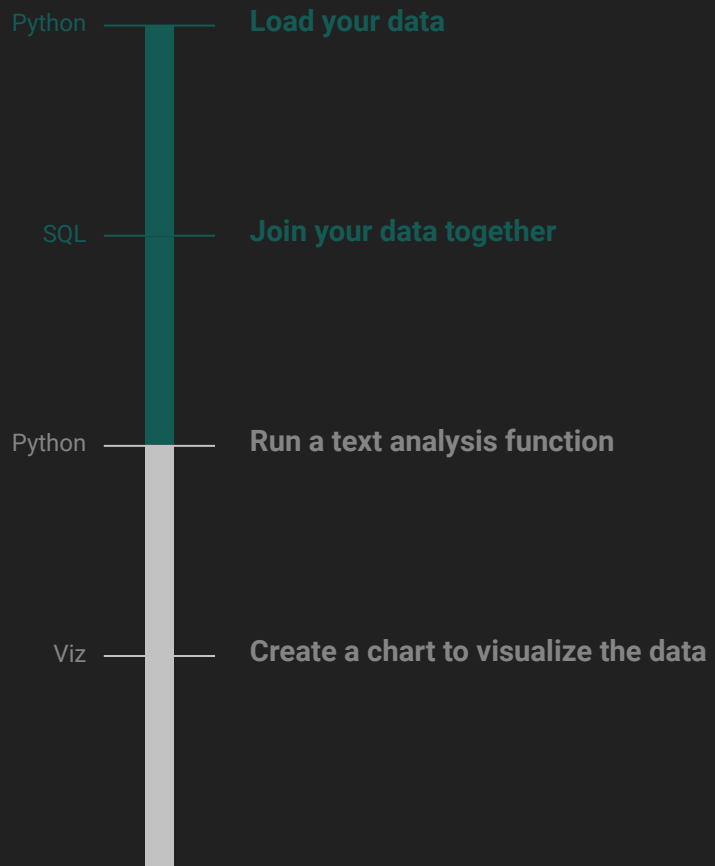
Agenda

- Designing data visualizations
- Dashboards vs. reports
- Interactive components
- Try out Hex’s visualization tools
- Data visualization process using python

We will use last week’s mock data analysis project, but don’t worry, this week’s template includes a copy of all the code we wrote last week!

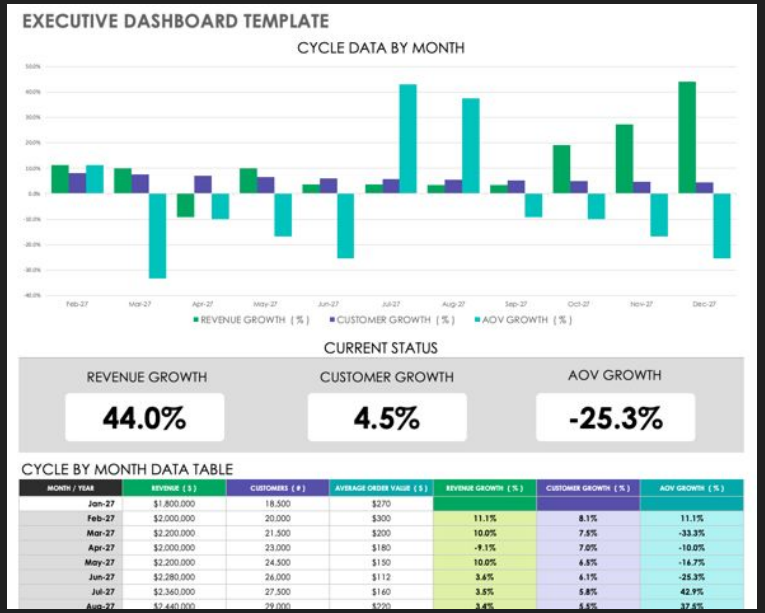
What is Hex?

- Hex is a SQL and Python notebook tool
- It works as a series of ‘cells’ that link together
- Here’s an example of some cells:

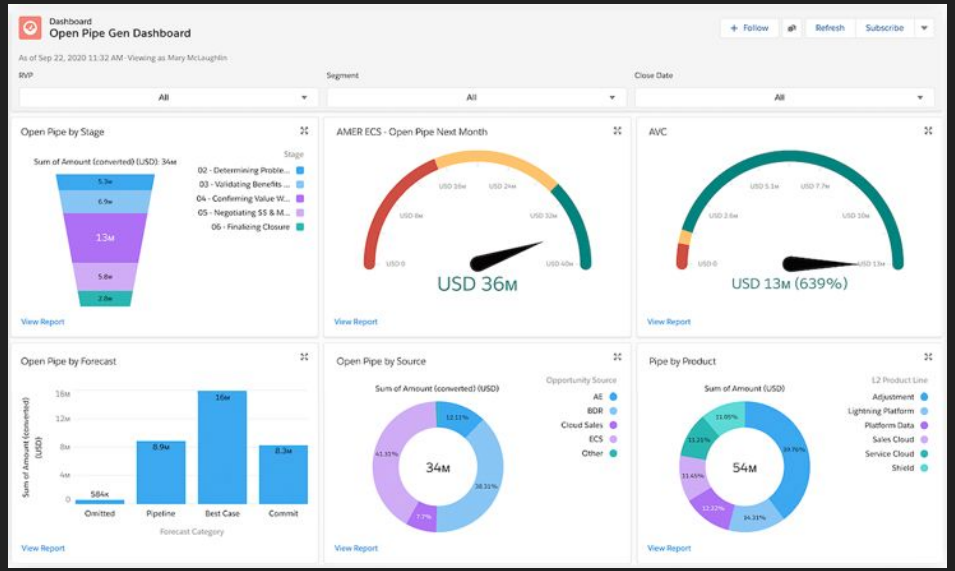


What is a dashboard?

- Tells a story to provide details, prompt action, or answer questions
- Often have some interactive components (filters, etc.)



MONTH / YEAR	REVENUE ($)	CUSTOMERS (#)	AVERAGE ORDER VALUE ($)	REVENUE GROWTH (%)	CUSTOMER GROWTH (%)	AOV GROWTH (%)
Jan-27	$1,800,000	18,500	$270			
Feb-27	$2,000,000	20,000	$300	11.1%	8.1%	11.1%
Mar-27	$2,200,000	21,500	$200	10.0%	7.5%	-33.3%
Apr-27	$2,000,000	23,000	$180	-9.1%	7.0%	-10.0%
May-27	$2,200,000	24,500	$150	10.0%	6.5%	-16.7%
Jun-27	$2,280,000	26,000	$112	3.6%	6.1%	-25.3%
Jul-27	$2,360,000	27,500	$160	3.5%	5.8%	42.9%
Aug-27	$2,440,000	29,000	$220	3.4%	5.5%	37.5%



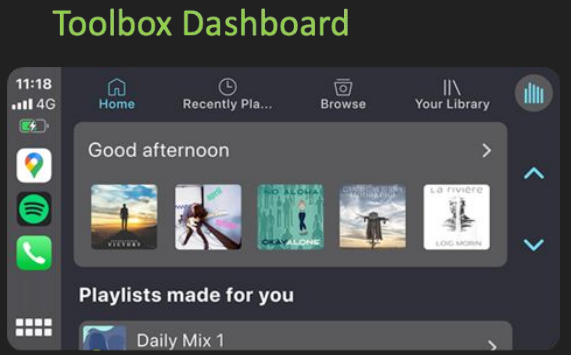
What is a report?

- Typically meant to provide a summary or status update, rather than prompt action
- Often includes written summary
- May be exported to PDF or Powerpoint

With modern visualization tools like Hex, Tableau, and PowerBI, the lines are blurring – reports can be interactive, and dashboards can include written context or links to other documents.

Dashboard styles

1. **Display Dashboards**: Summarize required info in a single view
2. **Action Dashboards**: Act as a ”to-do” list
3. **Toolbox Dashboards**: Give a set of filters and options for the user to explore

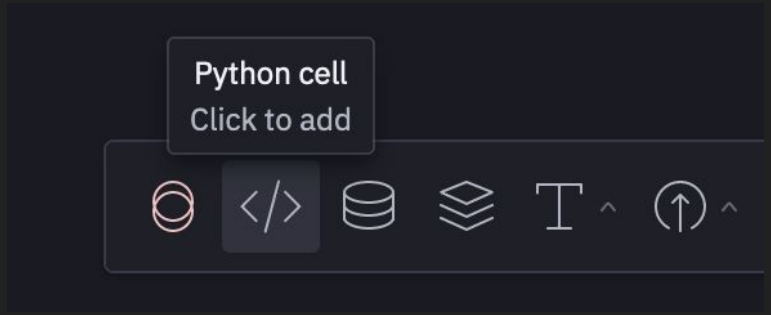


Develop your dashboard POV

- What question should this dashboard answer?
- What action should be taken by a viewer of this dashboard?
- What decisions should this dashboard inform?

Python and dataframes

- Add a new Python cell to your Hex project



- Check out one of the columns of your dataset:

  ratings_data[‘Title’]
- To run your code, click Run on that cell, or hit Command-Enter

Add another new Python cell

- Check out one of the rows of your dataset:

  ratings_data.iloc[0]

```
1  ratings_data['Title']
2
3  ratings_data.iloc[0]
```

Saving info to a dataframe

Similar to creating a variable, you can create a new column in any dataframe to save any results or calculations.

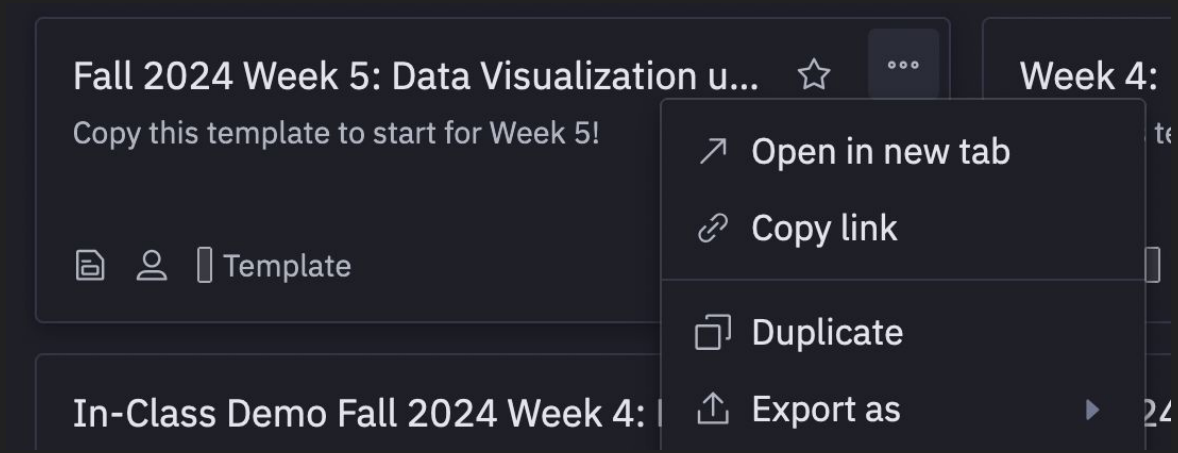
Example:

```
1  ratings_data['a_new_column'] = ratings_data['Season'] + ratings_data['Rating']
```

Then go look in your data browser (left side) under ratings_data!

Activity

1. Go to Projects, next to Fall 2024 Week 5 hit '...' then Duplicate



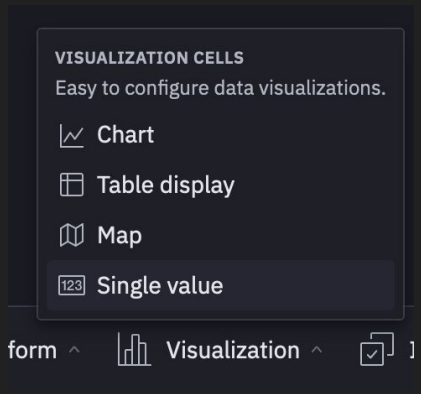
Building a single value into your dashboard

Some of the most powerful formulas for data analysis involve

We'll use python to summarize the dataframe and save to a variable:

```
cpv_average = campaigns2['cpv'].median()
```

And then add a single value cell:



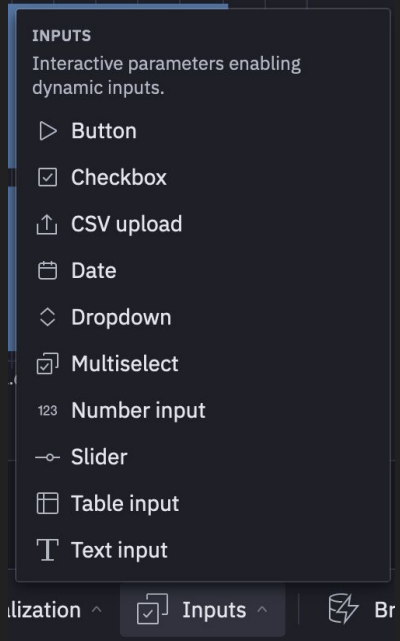
Iterating on your data as you visualize

As you create visualizations, you'll often realize that you need to go back to your data and make tweaks in SQL or python, and then go back and finish your visualization.

- What is your question?
- What is your desired output?
- What is your input?
- What is the first piece of info you need to answer the question?
- Is there an existing function that can help get you to your desired output?

Interactive components

Adding an interactive component to your dashboard helps your end user interact with the data! Think about what options they might like to filter the data by:



Layout

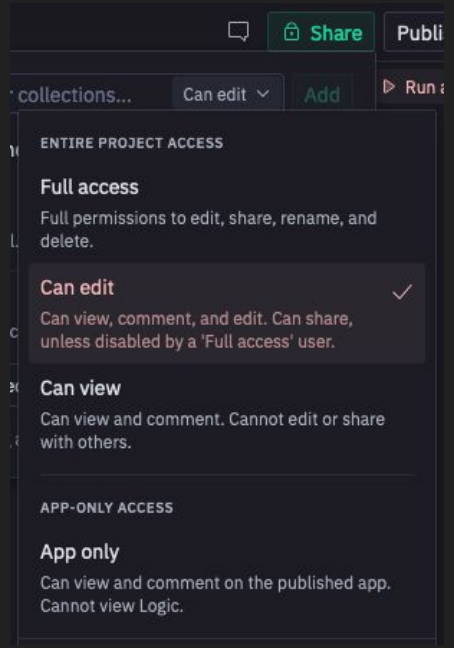
When laying out your dashboard in App view, in general you want to put the most summarized view at the top, and more detailed views below

- Context is key! Add Text blocks (type Markdown or Text) to add headers, sections, or descriptive info
- Line up related information into the same column or row, so that the eye can visually make the connection
- Avoid blank space – make cells large enough to take up the full width of the screen when feasible
- Order cells logically for the viewer – they can be in a different order than in the notebook view

Publishing and sharing

With any visualization tool, there are publishing or sharing settings to consider

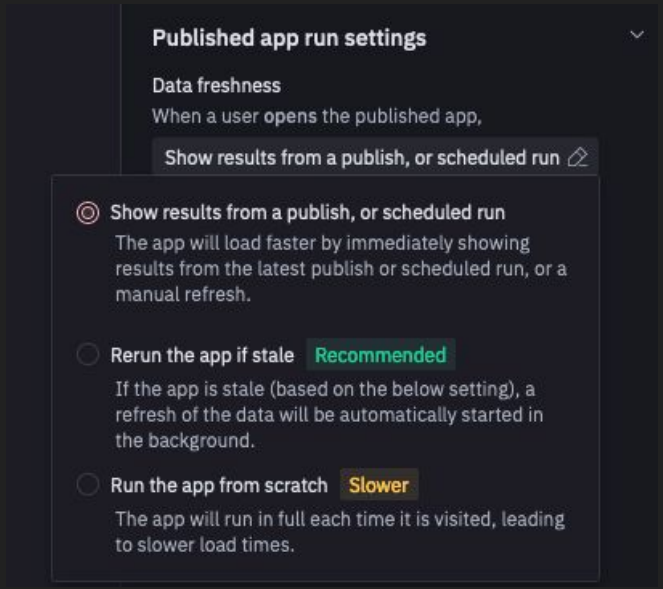
- In Hex, your ‘Published’ version can be different than your draft version, if you make more edits after you hit Publish
- Then you’ll need to set up sharing:
  - For non-technical users, “App only” is best
  - For collaborators, “Can edit” is usually best
- Privacy vs. transparency



Refreshing the data

Each time you publish, check that the data refresh settings make sense

- Dashboards should not refresh more often than your data source itself
- If your data is static → results from publish
- If refreshing from a database → rerun if stale



Review

- Designing your visualization (dashboard, report, etc.)
- Setting your Point of View
- Charts and tables
- Interactive components
- Layout considerations
- Publishing and sharing
- Refreshing the data

Thanks so much for attending!

You have access to your account on the Hex Professional plan for the year! You can use it for practice or projects.

- PolicyViz blog
- Data modeling: docs.getdbt.com/best-practices/how-we-style/1-how-we-style-our-dbt-models
- learn.hex.tech
- Learn SQL via a Murder Mystery: https://mystery.knightlab.com/

And, feel free to keep in touch via LinkedIn!