Programming, Problem Solving, and Algorithms

CPSC203, 2021 W2

Announcements

- **Test 2 is this week**
- **Lab3 is this week**
  - It's about DataClasses!
- **Problem of the Week 3 is due this week (extended from last week)**
  - Assigned Pandas videos from Thursday should help!
- **Problem of the Week 4 is also due this week**
  - Dataclasses practice

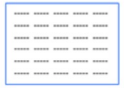
Today's Plan...

1. Announcements! (5 mins)
2. Review/Questions (10 mins)
3. Billboard 100 Demo (65 mins)

Slides from the Assigned Videos

103 to 203

Typical Introductory Data Flow:

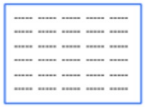


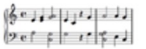
.csv file

Python problem solution using simple data types and elementary list iteration.

Matplotlib bar or line graph or other summative output illustrating results of computation.

CPSC103++ Data Flow:

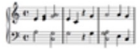




Diverse data sources

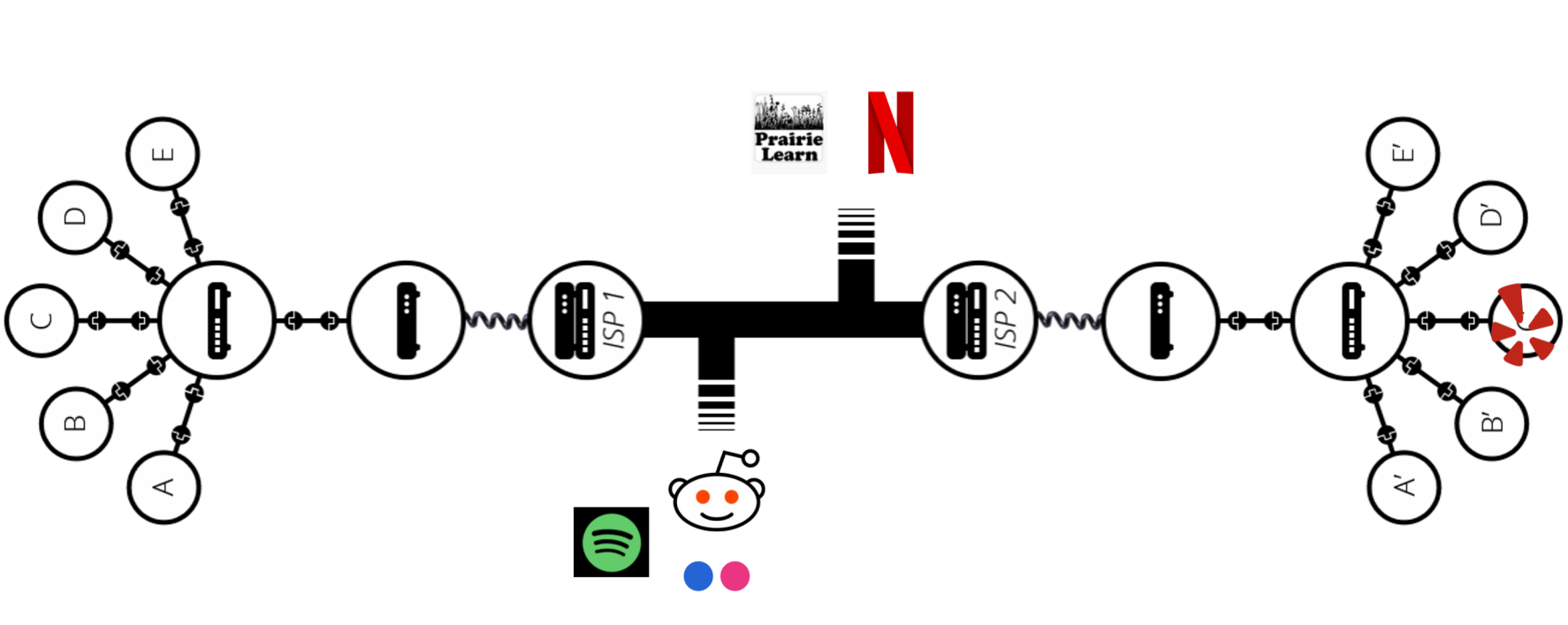
data synthesis

Analysis and data assembly



Diverse outputs

The internet...



Billboard Hot 100...

Navigate to https://www.billboard.com/charts/hot-100

What happens to the URL if you load a past week?_____

What happens to the page if you substitute a different date into the URL?

_____

_____

Write one question you would like to ask of this data: _____

_____

_

Anatomy of html...

```
<!DOCTYPE html>

<html><head><title>The Dormouse's story</title></head>

    <body><p class="title"><b>The Dormouse's story</b></p>

        <p class="story">Once upon a time there were two little sisters.
    Their names were <a href="http://example.com/elsie" class="sister"
    id="link1">Elsie</a>, and <a href="http://example.com/lacie"
    class="sister" id="link2">Lacie</a>, and they lived at the bottom of
    a well.</p>

    </body>

</html>
```

Billboard Hot 100... page source

    49   <img src="https://assets.billboard.com/assets/1568911107/images/charts/arrow-down.svg?df89925e3b37f64521bd" srcset="https://assets.billboard.com/assets/1568911107/images/charts/arrow-down-mobile.svg?df89925e3b37f64521bd 30w, https://assets.billboard.com/assets/1568911107/images/charts/arrow-down.svg?df89925e3b37f64521bd 38w" sizes="(min-width: 768px) 38px, 30px">

<img src=" gAAAABJRU5ErkJggg==" data-src="https://charts-static.billboard.com/img/2019/08/taylor-swift-p7u-lover-tuk-53x53.jpg" data-srcset="https://charts-static.billboard.com/img/2019/08/taylor-swift-p7u-lover-tuk-53x53.jpg 53w, https://charts-static.billboard.com/img/2019/08/taylor-swift-p7u-lover-tuk-106x106.jpg 106w, https://charts-static.billboard.com/img/2019/08/taylor-swift-p7u-lover-tuk-87x87.jpg 87w, https://charts-static.billboard.com/img/2019/08/taylor-swift-p7u-lover-tuk-174x174.jpg 174w" sizes="(max-width: 767px) 72px, (min-width: 768px) 86px" class="chart-list-item__image" alt="Taylor Swift Lover Billboard Hot 100">

Lover

Lover

Taylor Swift

Song Lyrics

23
LAST WEEK

10
TWO WEEKS AGO

10
PEAK POSITION

4
WEEKS ON CHART

Beautiful Soup

Reads the html source into a data structure that's easy to query!

https://www.crummy.com/software/BeautifulSoup/bs4/doc/

```
html = simple_get("https://www.billboard.com/charts/hot-100" + '/' + date)
mydivs = html.findAll("div", {"class": "chart-list-item"}) // all the data is here!!

for div in mydivs:
    s = Song(div.attrs['data-title'], div.attrs['data-artist'], int(div.attrs['data-rank']))
```

Still too messy for us! Remedy? https://github.com/guoguo12/billboard-charts

demo...

Some challenges...

Given last week's chart,

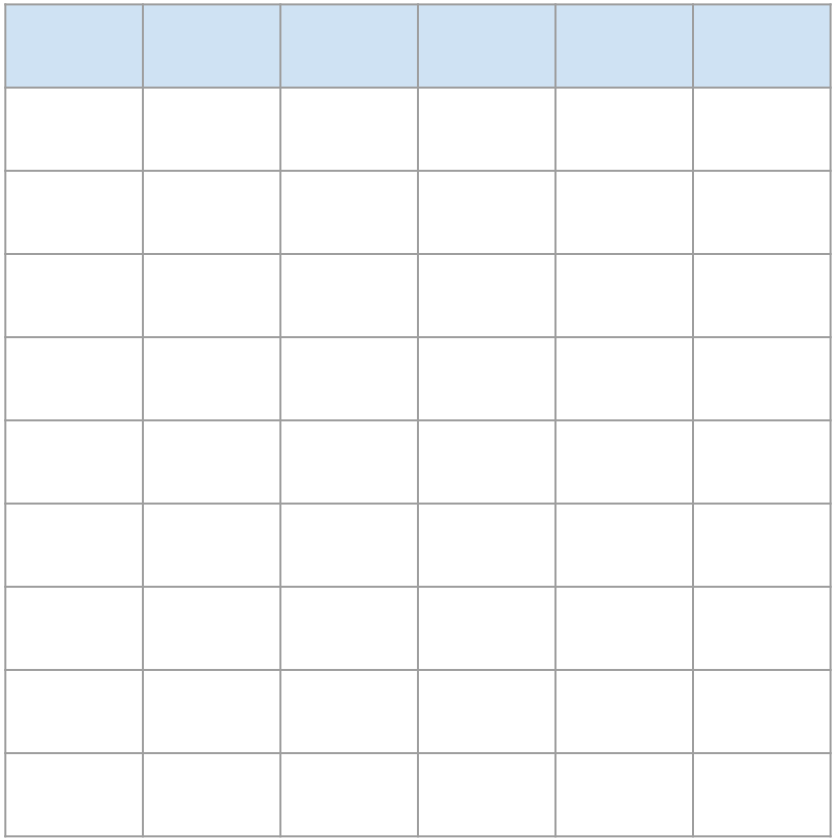
1) How many new songs were there?
2) What's the average peak?
3) Among those who were on the list for more than 10wk, what's the average peak? (is it very different than the previous answer?)
4) Which song changed the most? Was it rising or falling?
5) Write and answer your own question:

_____

Some challenges...

Given last week's chart,

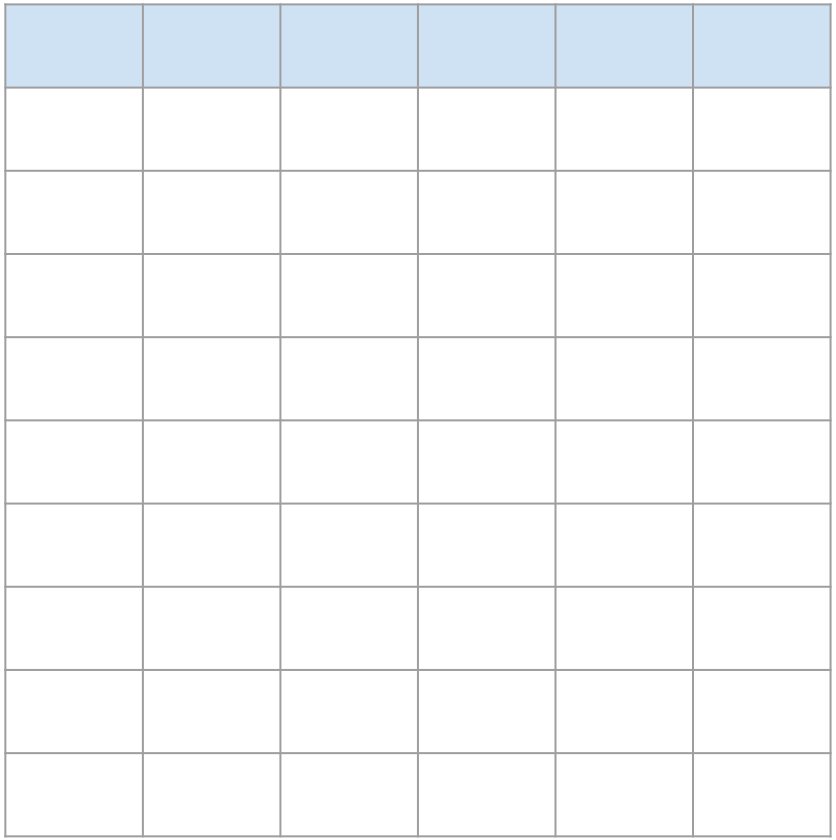
How many new songs were there?



Some challenges...

Given last week's chart,

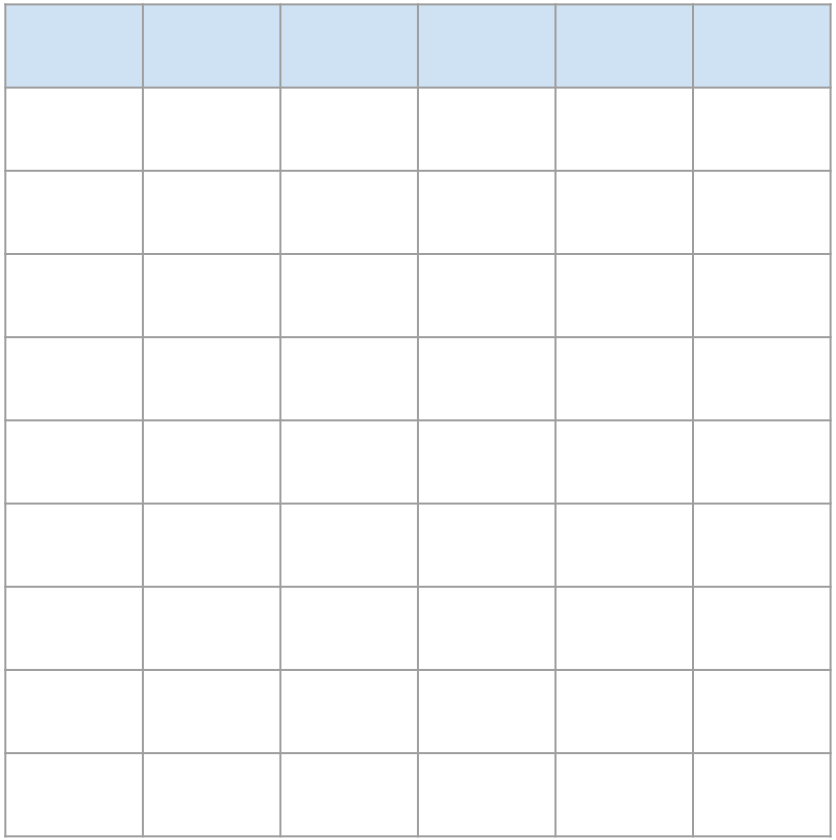
What's the average peak?



Some challenges...

Given last week's chart,

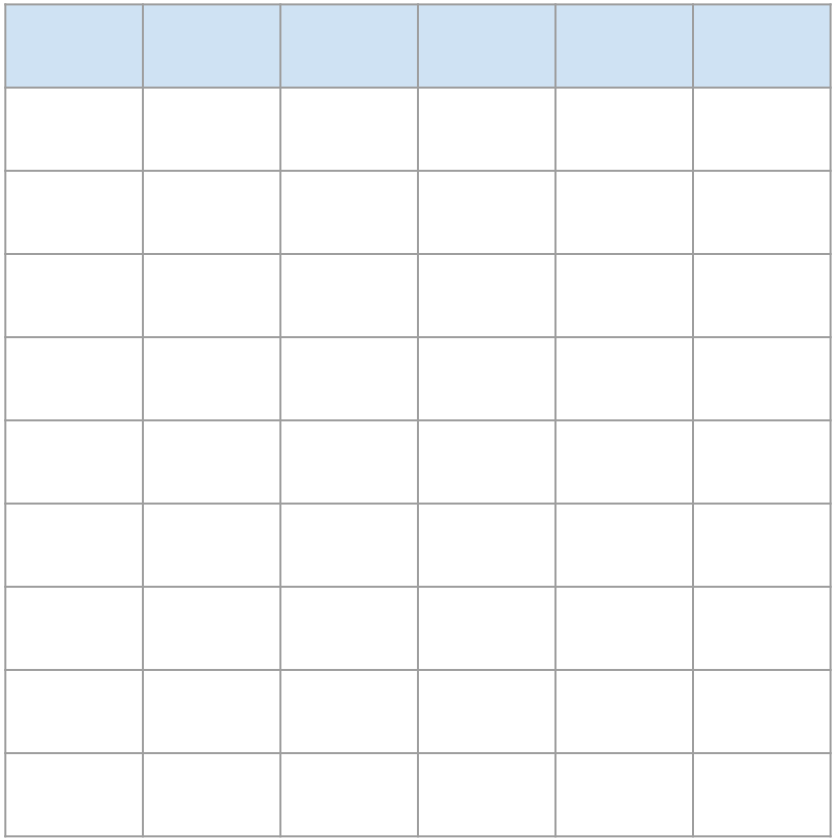
Among those who were on the list for more than 10wk, what's the average peak? (is it very different than the previous answer?)



Some challenges...

Given last week's chart,

Which song moved the most? Did it rise or fall?



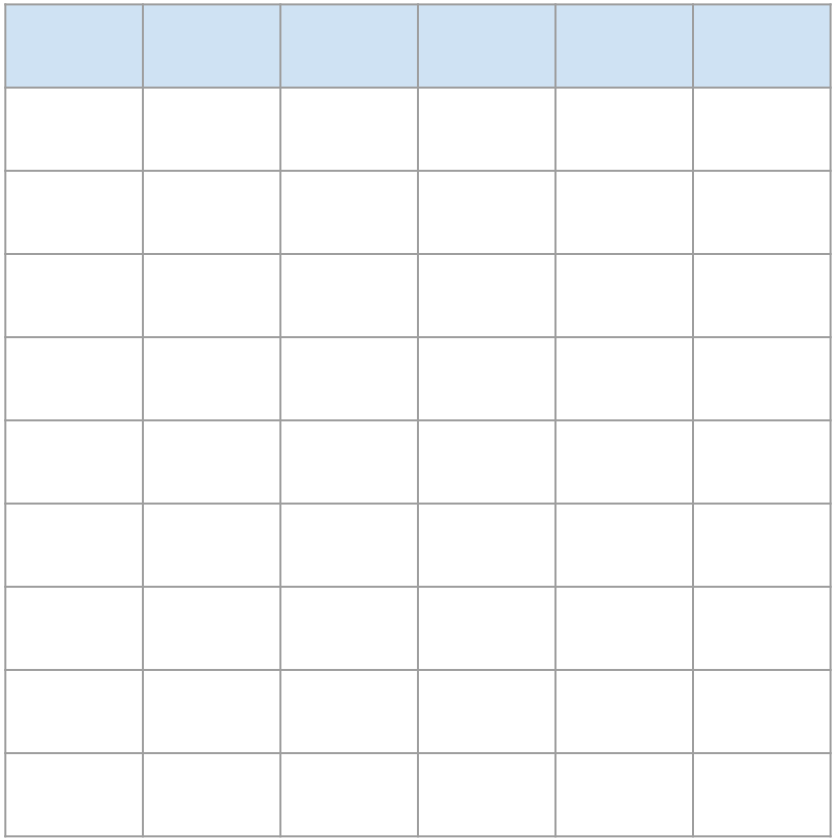
Some challenges...

Given last week's chart,

Write and answer your own question:

_____

_____



ToDo for next class...

POTD: Continue every weekday! Submit to PL.

References:

https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf

https://www.crummy.com/software/BeautifulSoup/bs4/doc/