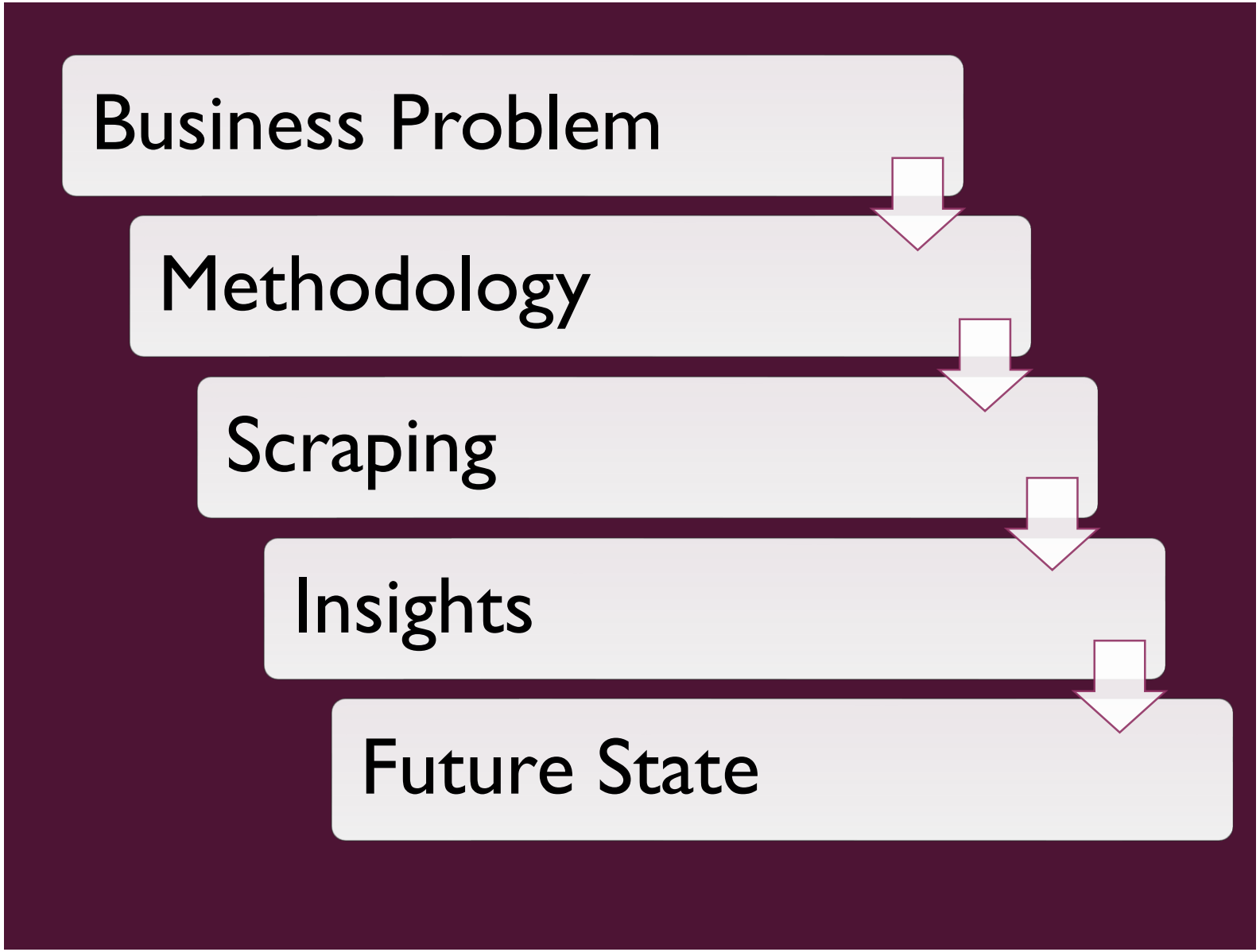
AXS EVENT INSIGHTS

KWEKU ULZEN

NYC DATA SCIENCE ACADEMY: COHORT #12

AGENDA


Business Problem
Methodology
Scraping
Insights
Future State

ABOUT AXS

- AXS is a digital marketing platform for purchasing tickets for sports and entertainment events in the US, and overseas.
- Partner with promoters and venues to sell tickets directly to consumers
- Large customer base in New York, Los Angeles, Atlanta, and London
- Clearly dislike people scraping their website

BUSINESS PROBLEMS

- Book artists to perform in front of their optimal audience
  - Appropriately price venue and show
  - Schedule performer at the right time of week
  - Choose a venue of adequate size
  - Identify new growth opportunities

METHODOLOGY

1. Identify metrics to solve business problems
2. Scrape AXS Events
3. Clean Data
4. Visualize Insights

axs Sign in HEADLINES MUSIC ARTS & FAMILY SPORTS NEW YORK CITY , N

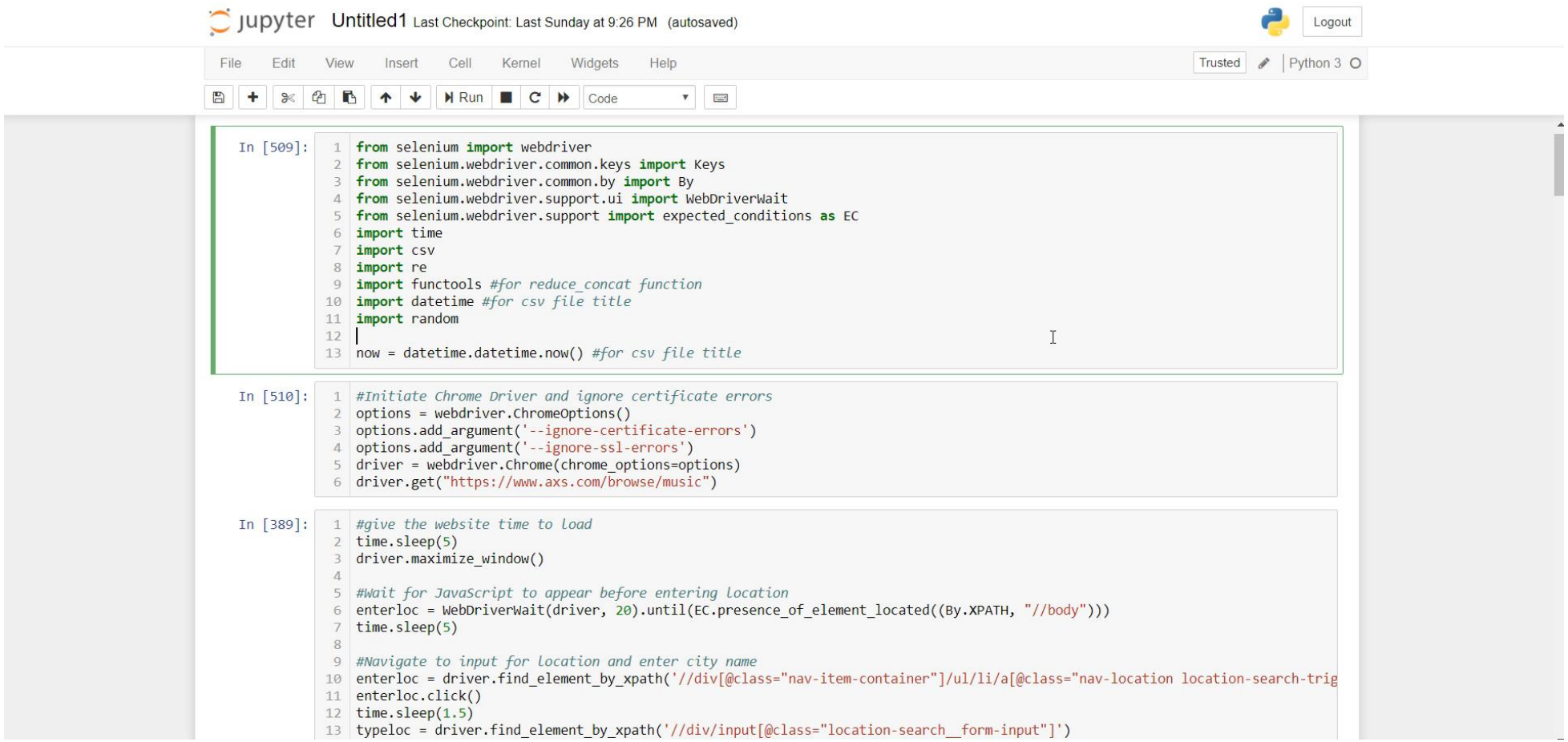
MUSIC

All event results matching ""

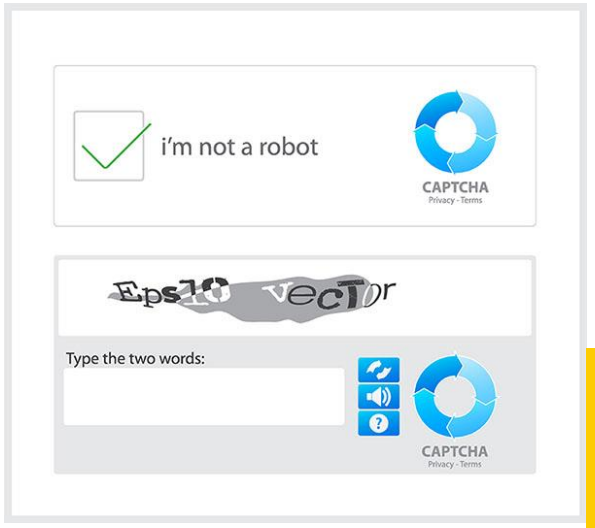
Date	Event
FEB 14 WED	Robert Plant & The Sensational Space Shifters Beacon Theatre, New York City NY
FEB 14 WED	Village of Love Music Hall of Williamsburg, Brooklyn NY
FEB 15 THU	August Burns Red Starland Ballroom, Sayreville NJ
FEB 15 THU	An Evening with Noel Gallagher's High Flying Birds Radio City Music Hall, New York City NY
FEB 15 THU	BØRNS Brooklyn Steel, Brooklyn NY
FEB 15	Dirty Fences Rough Trade NYC, Brooklyn NY

SCRAPING

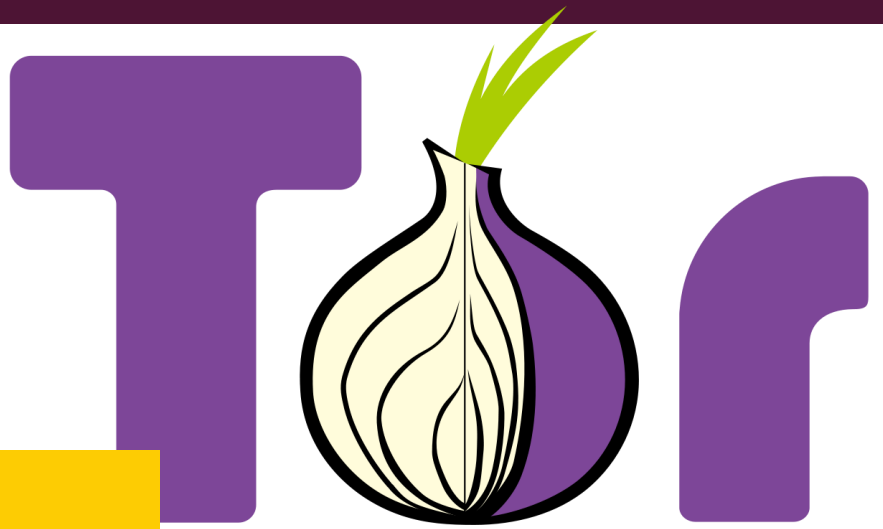
SCRAPING


jupyter Untitled1 Last Checkpoint: Last Sunday at 9:26 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [509]:
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import csv
import re
import functools #for reduce_concat function
import datetime #for csv file title
import random
now = datetime.datetime.now() #for csv file title
In [510]:
#Initiate Chrome Driver and ignore certificate errors
options = webdriver.ChromeOptions()
options.add_argument('--ignore-certificate-errors')
options.add_argument('--ignore-ssl-errors')
driver = webdriver.Chrome(chrome_options=options)
driver.get("https://www.axs.com/browse/music")
In [389]:
#give the website time to load
time.sleep(5)
driver.maximize_window()
#Wait for JavaScript to appear before entering location
enterloc = WebDriverWait(driver, 20).until(EC.presence_of_element_located((By.XPATH, "//body")))
time.sleep(5)
#Navigate to input for location and enter city name
enterloc = driver.find_element_by_xpath('//div[@class="nav-item-container"]/ul/li/a[@class="nav-location location-search-trig
enterloc.click()
time.sleep(1.5)
typeloc = driver.find_element_by_xpath('//div/input[@class="location-search__form-input"]')

SCRAPING


i'm not a robot
CAPTCHA
Privacy - Terms
Eps10 vector
Type the two words:
CAPTCHA
Privacy - Terms

HMA!


Tor

Se

SCRAPING

jupyter Untitled1 Last Checkpoint: Last Sunday at 9:26 PM (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

```
typeloc.send_keys(i)
time.sleep(0.1)
typeloc.send_keys(Keys.END)
time.sleep(0.1)
time.sleep(2)
typeloc.send_keys(Keys.ENTER)
time.sleep(3)
```

In [196]:

```
#scroll to end of page so that all elements are visible
calm = WebDriverWait(driver, 20).until(EC.presence_of_element_located((By.XPATH, "//html"))) #if elements aren't available it
time.sleep(3)
lenOfPage = driver.execute_script("window.scrollTo(0, document.body.scrollHeight);var lenOfPage=document.body.scrollHeight;re
match=False

while(match==False):
    lastCount = lenOfPage
    time.sleep(3.5)
    lenOfPage = driver.execute_script("window.scrollTo(0, document.body.scrollHeight);var lenOfPage=document.body.scrollHeigh
    if lastCount==lenOfPage:
        match=True

scroll = driver.find_element_by_xpath('//html')
time.sleep(2)
scroll.send_keys(Keys.HOME)
time.sleep(2)
```

In [236]:

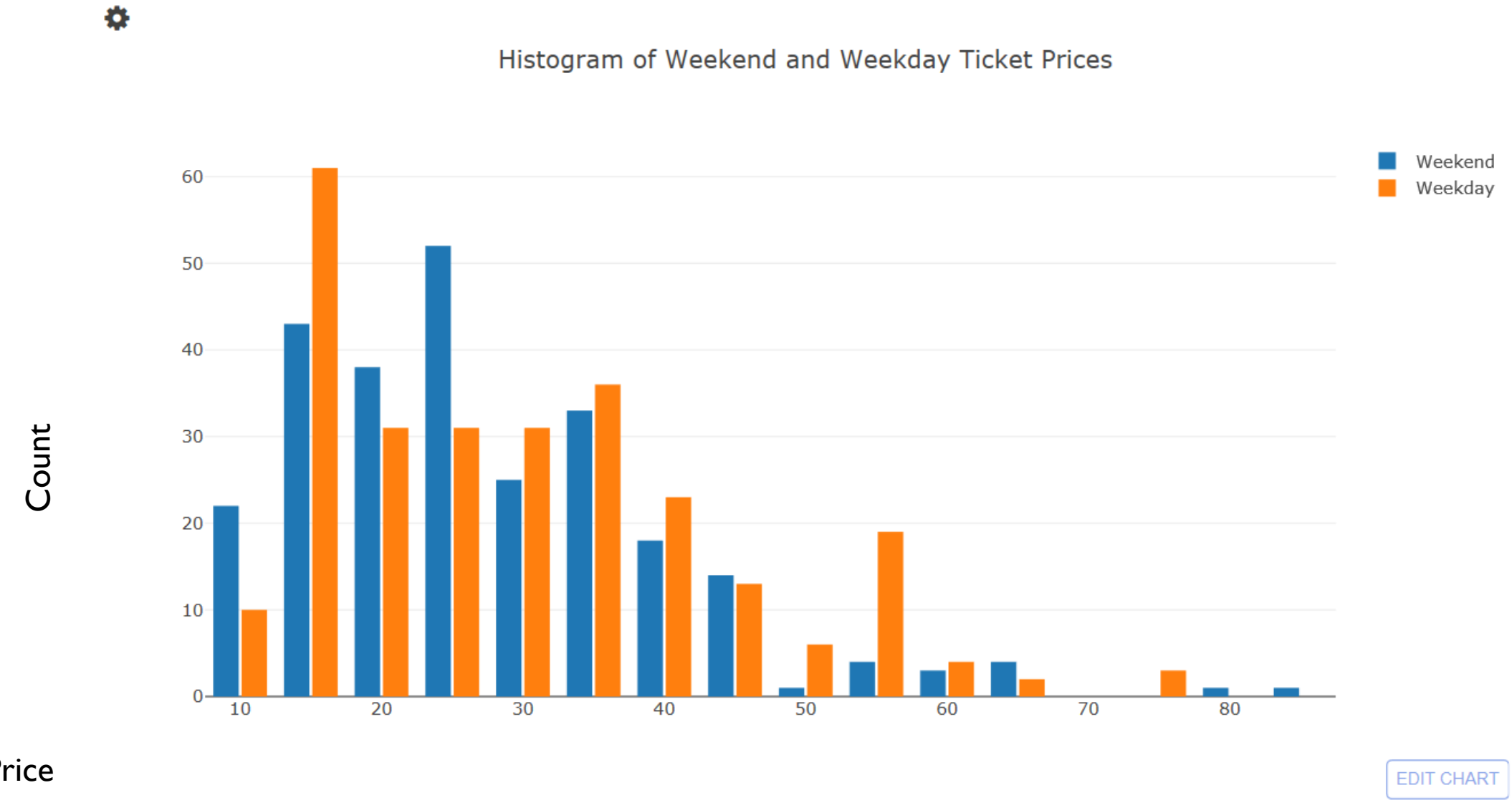
```
#Open CSV Writer
filename = paste(['axs_events_'], [now.year], ['-'], [now.month], ['-'], [now.day],['.csv'], sep='')
filename = ''.join(filename)
csv_file = open(filename, 'a', newline = '')
writer = csv.writer(csv_file)

#collect html tags, this will need to go inside of the nested loop function
```

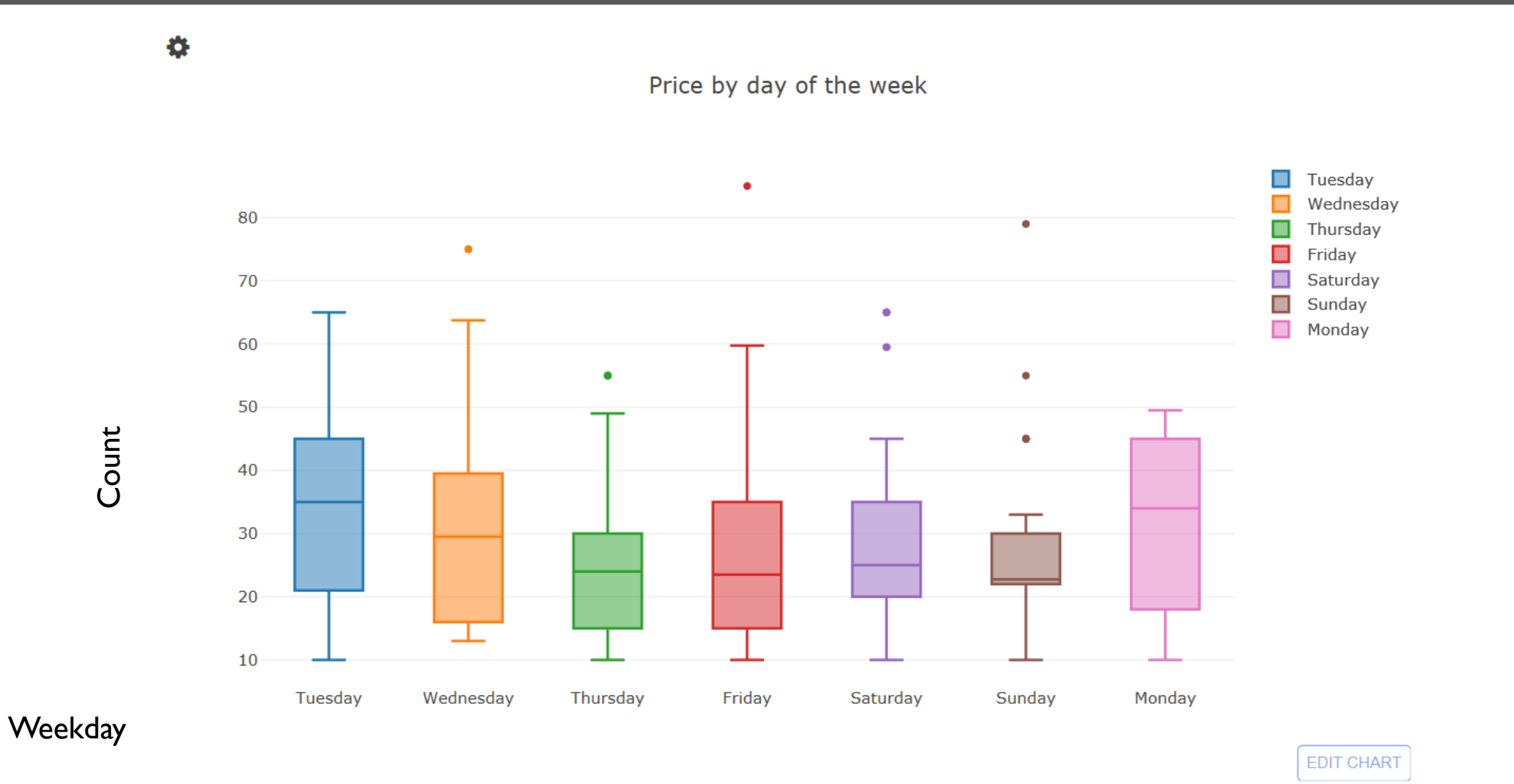
DATA CLEANSING TASKS

Tasks	Action
Combine dates	Apply to_datetime function to combine independent month day and year features; create Boolean weekend column
Remove duplicates	*df.drop_duplicates()*
Remove commas from venue	Regex replace ‘,’ with ‘’ on desired column
Split performers on key words	Regex split on ‘with’ | ‘*:’ | ‘and’ in a reversed list
Remove extra characters from time	Split on ‘- ‘ to remove city from time
Convert string time to a time type	Apply to_datetime function
Remove extra characters from price	Regex split on ‘ – ’ | ‘starting from: ’ to eliminate non-numeric characters
Split price into low and high column	Create new columns in dataframe for items in columns created from split
Reorder columns	Send column names to list and manipulate
Visualize	Plotly

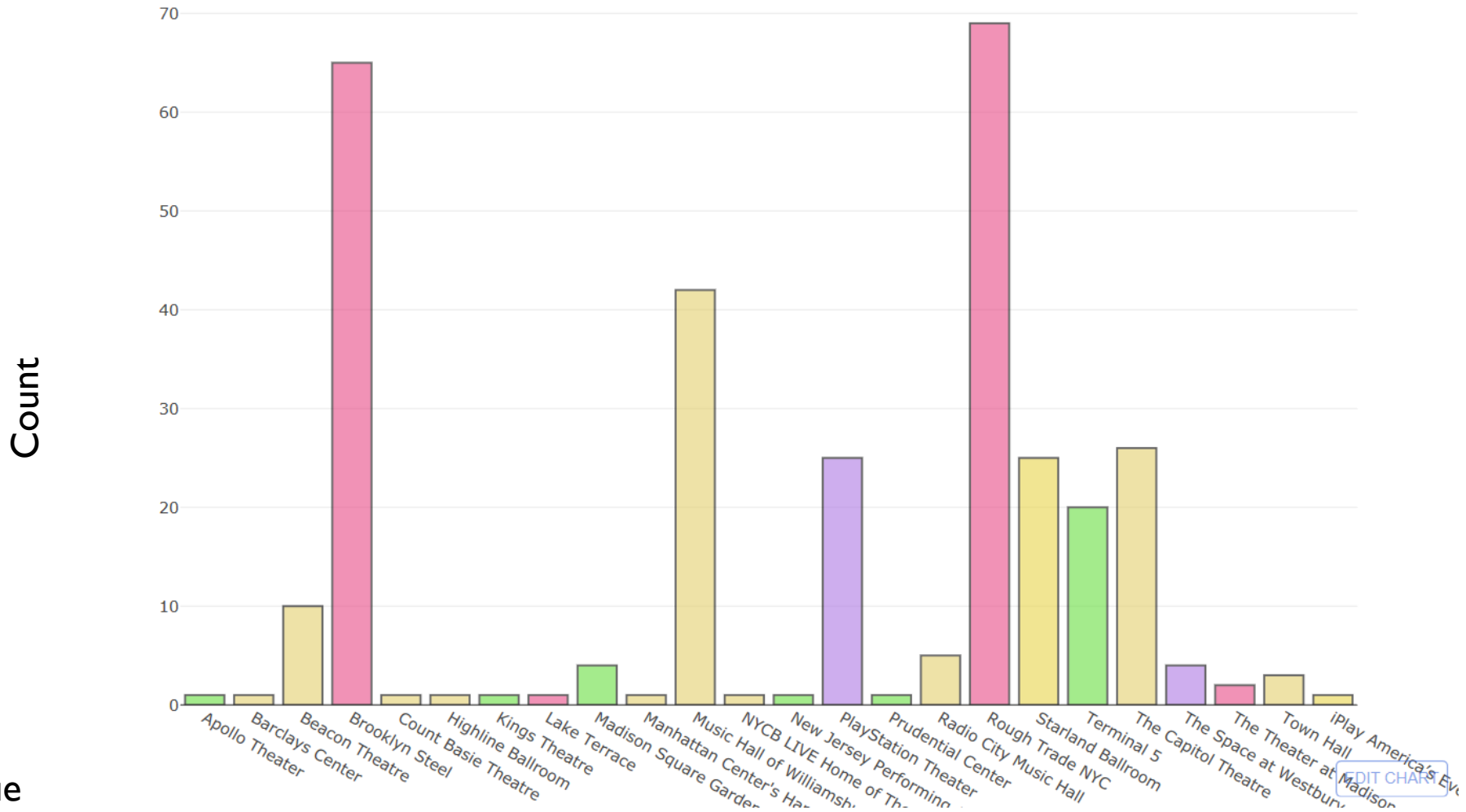
INSIGHTS

- Ticket Prices
- Best weekdays for events
- Well represented venues
- Artist concert opportunities

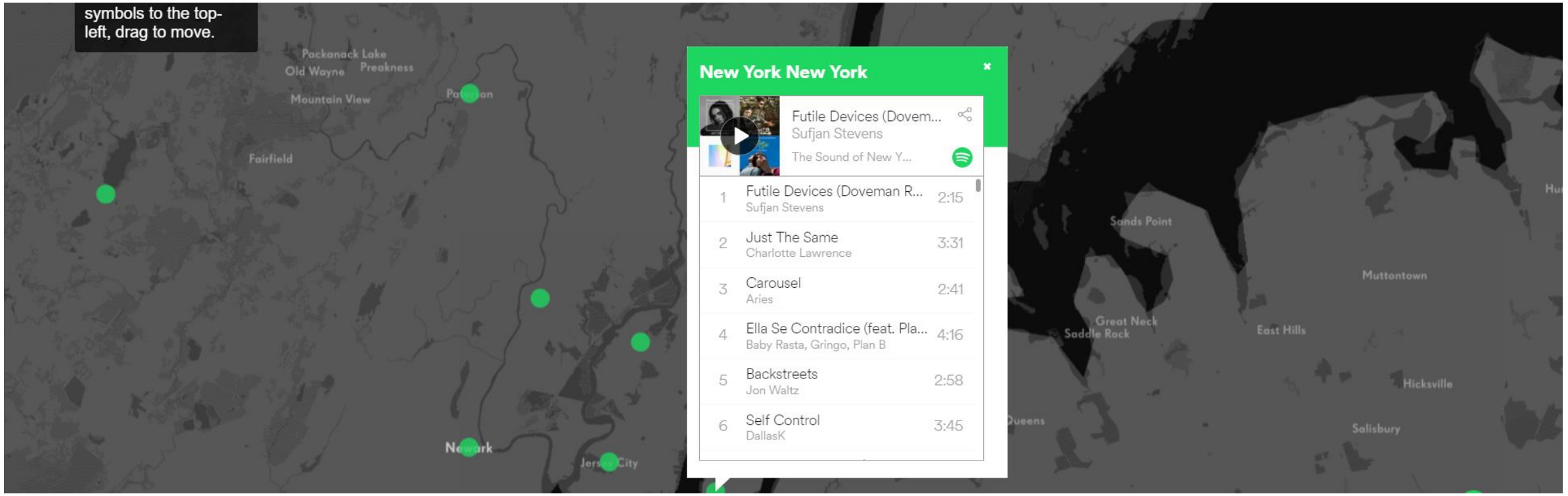

Histogram of Weekend and Weekday Ticket Prices
Weekend
Weekday
Count
Price
0
10
20
30
40
50
60
10
20
30
40
50
60
70
80
EDIT CHART


Price by day of the week
Count
80
70
60
50
40
30
20
10
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Monday
Weekday
EDIT CHART

Events by Venue



Venue



OPPORTUNITIES

FUTURE STATE

Organize	Organize data for consumers to more easily plan events and purchases
Map	Map locations of concerts
Incorporate	Incorporate Spotify API to alert consumers of similar concerts
Combine	Combine with larger ticket sites to have a larger sample of events
Perform	Perform analysis for sporting events, comedy, and theatre
Collect	Continuously collect information to predict future occurrences against historical data