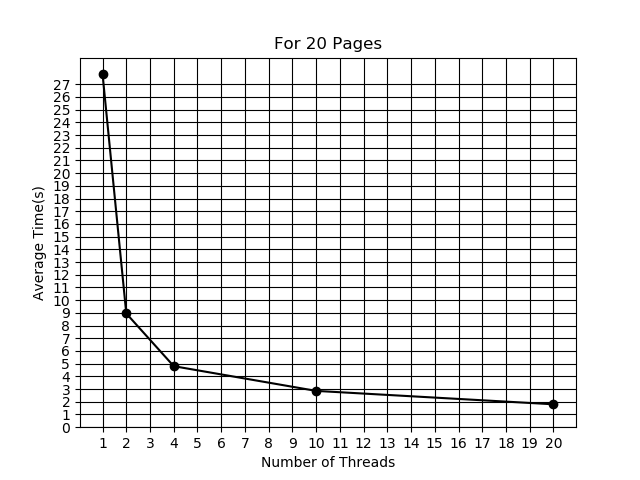
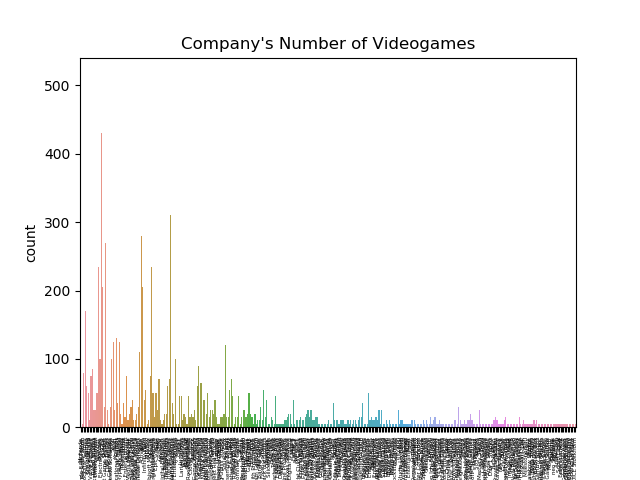
EECS 338 Final Project Design Document

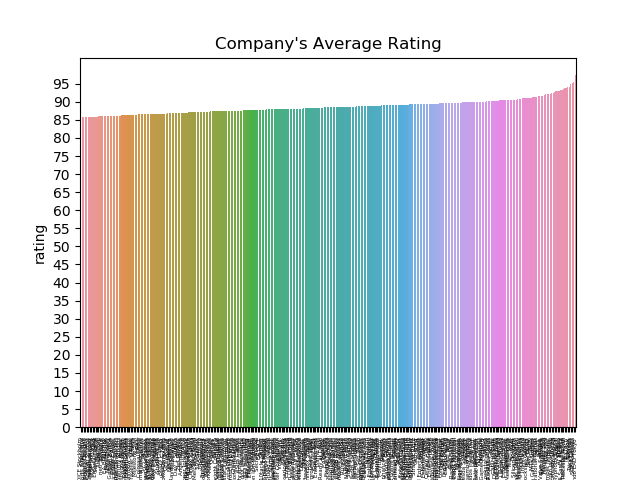
* Files Included
  + Scraper.py: Takes in number of threads to be created and number of pages to be scraped and creates that number of threads in order to scrape that many pages. (Number of pages must be divisible by number of threads; otherwise, we raise an error). Times the process and writes it to txt file. Writes all the data gathered (Videogame name, Rating, Company Name, and Year Created) to both a csv and to individual txt files.
  + Main.py: Takes in 5 command line arguments (nthread1, nthread2, nthread3, nthread4, nthread5, npages). Will then run 5 scrapes with the number of threads specified. Will plot the number of threads versus the average time needed to scrape the website by opening the time.txt file for each thread number and calculating the average. Will call Plot to plot the rest of the data.
  + Plot.py: Will plot the average rating of the videogames per company as well as another graph that portrays how many videogames that company has released that is within the range of pages we scraped. Needs the gamesdata.csv as input.
* Major Data Structures:
  + Threads: created using myThread class in the Scraper file. The number of threads used is specified by the main method.
  + Arrays: Global arrays which store all data are created in the Scraper file. They keep track of all the relevant data (videogame name, rating, company name, year created) and have the threads all add to them so we can process the data later.
  + Panda’s Dataframe: Stores all the data. This is created in the Scraper file and is acted on in the Plot file.
* Sample Output
  + Sample Graphs:



Graph 1: Scraping 20 Pages with various number of threads vs average time to scrape



Graph 2: The company vs the amount of videogames it has in the pages we scraped



Graph 3: company vs average rating their games achieved

* Sample txt output files:
  + Rating:
    - 97.64

97.54

97.42

97.35

97.33

97.04

97.01

96.67

96.56

96.43

* Name
  + Super Mario Galaxy

The Legend of Zelda: Ocarina of Time

Super Mario Odyssey

Super Mario Galaxy 2

The Legend of Zelda: Breath of the Wild

Grand Theft Auto IV

Grand Theft Auto V

Grand Theft Auto IV

SoulCalibur

* Date
  + 2007

1998

2017

2010

2017

2008

2013

2008

1999

2009

1996

* Company
  + BioWare/Electronic Arts

Naughty Dog/SCEA

Bungie/Microsoft Game Studios

Valve Software/VU Games

DMA Design/Capcom

Valve Software

Bethesda Game Studios/Bethesda Softworks

Naughty Dog/SCEE

KCEJ/Konami

* Time
  + Thread 1: 1.83

Thread 2: 1.97

Thread 3: 1.93

Thread 4: 1.87

Thread 5: 1.68

Thread 6: 1.76

Thread 7: 1.6

Thread 8: 1.8

Thread 9: 1.89

Thread 10: 1.89

* Gamedata.csv

Rockstar North/Rockstar Games,Grand Theft Auto III,93.54,2002

Infinity Ward/Activision,Call of Duty 4: Modern Warfare,93.54,2007

Kojima Productions/Konami,Metal Gear Solid 4: Guns of the Patriots,93.53,2008

Bungie/Microsoft Game Studios,Halo 3,93.53,2007

Neversoft Entertainment/Activision,Tony Hawk's Pro Skater 3,93.43,2001

Infinity Ward/Activision,Call of Duty: Modern Warfare 2,93.42,2009

Atlus,Persona 5,93.36,2017