Zoe Block

INFO 664

I’ve found the final project for this course to be extremely informative, a great learning experience, and remarkably frustrating. The frustration taught me some hard lessons about wrangling Python, and some meaningful ones about myself as well. I embarked on this project with lofty ambitions. Over the course of several semesters, I’ve seen classmates produce beautiful visualizations in Tableau and complex webpages featuring interactive widgets, and I envied their capabilities every time. This project felt like the perfect opportunity to acquire some of these skills for myself.

Once I found an appealing dataset—a process that required two tries, which in retrospect feels like an omen of the challenges ahead—I decided that I would use the data to produce basic visualizations in Tableau and record my findings on a website hosted on Glitch. Neither of these deliverable goals came to be, in the end. There were two reasons for this: the first was the technical obstacle of the sheer size of my dataset of choice, and the second was my aspiration to deliver something at a skill level I couldn’t realistically have ever produced. I have extensive XML experience from my career in publishing, producing eBooks and online content and designing XML-based digital workflows. Working with XML has led me to a decent grasp of XPath, a very rudimentary sense of Regex, and some confidence with HTML and CSS. None of these things translates into a natural facility with Python, but there I was, overcome by the talents of my classmates and abandoning common sense in favor of output I thought would be impressive or cool.

I don’t want to minimize the technical difficulties I experienced. I’d chosen a dataset of 19m board game reviews, with an additional dataset of 21k games and associated metadata. I went through Kaggle (the source of the datasets), Jupyter Notebook, and finally Jupyter Labs as workspaces, and only Jupyter Labs (hosted on Google Cloud AI’s notebook) crashed occasionally rather than regularly. Most of the Python I’ve produced is for narrowing the dataset to a workable size rather than actually analyzing the reviews and finding information. The timeliness of the project and my research question (whether there was a relationship between the onset of Covid-19 and game usage, and what kind of games were people playing in lockdown) was one of the things that made the dataset appealing. The few results I managed to produce were incomplete—I identified 30 reviews that referenced Covid-19, but I struggled with linking the reviews with their corresponding game names and couldn’t manage to de-dupe the game name list.

I’m still inspired by the idea behind my project and want to continue to work on it, as time allows. The skills I learned in this class are a great foundation for producing the kind of elaborate results I wish I could do. But the project overall has been a powerful reminder that it’s okay to focus on building that solid foundation, and that it isn’t possible to leapfrog the basics to get to the end. It’s something I will have to remember in the future, whether working with Python or anything else.