Switching Topics

By Ricky Gajeski
Cheyann Odle

Revising Topic

Original Dataset: College Majors and Job salaries

Why?: Data set originally found did not fit the scope of our project and were unsure about where to find datasets that would be specific without having to

New Dataset: Video Game Sales

New Question: What are the best variables to predict high sales for a video game?

Dataset: title, years, company, genre, console, NA, EU, JA, Global sales

Progress

- So far, we have outlined our new schedule and have started working on the more basic data analysis, while coming up with unique ways to compare our data (goals).

Goals:

- Finish Cleaning
- Check outlier years
- Do correlational analysis on a few variables

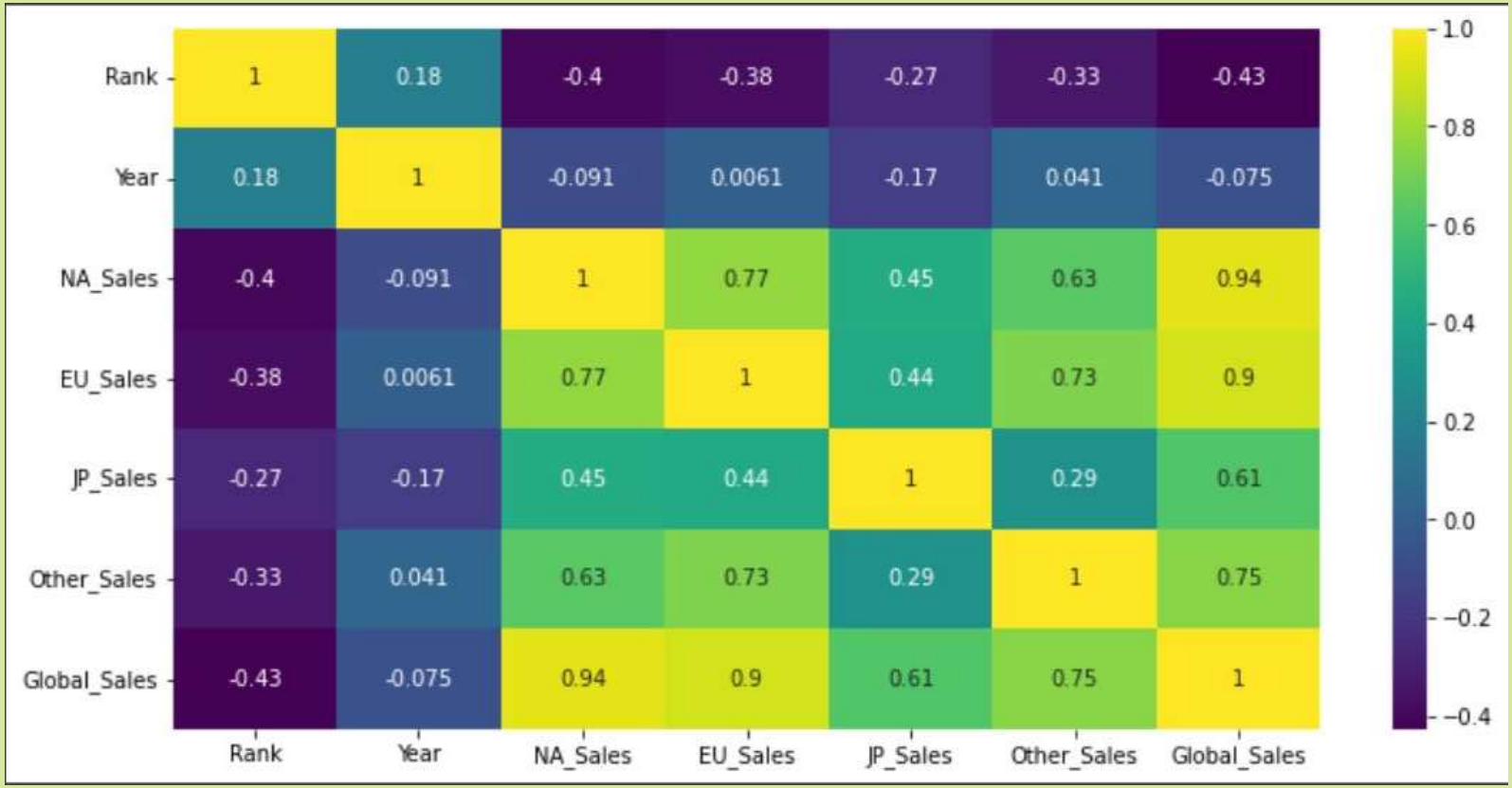
Challenges and Possible Counters

- Writing the code
  - Asking people, past labs
- Google Colab
  - Pro: can work in one place & familiar
  - Con: doesn't work with both people on it at once to well
- Deciding what direction we want to exactly take this project
  - Communication, spitballing and weighing ideas (is it possible)
  - Goal list at the beginning of CoLab page (initial when done)
- Balancing Data Project and other Assignments
  - Switching from working on and off by ourselves and informing the other to setting some time on Tuesday for specifically the data project

General 5 Number Summary

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000
mean	8290.190228	2006.405561	0.265647	0.147731	0.078833	0.048426	0.540910
std	4792.654450	5.832412	0.822432	0.509303	0.311879	0.190083	1.567345
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4132.500000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8292.000000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12439.500000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.480000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

Charts and Analysis



Predictive Modeling

Logistic regression: doesn’t need variables to be linear to relate

New Schedule

Schedule:

- **March 24th:**Ricky & Cheyann Presentation/ Preliminary/ Data

***Start meeting on Tuesdays**

- **April 19th:** Ricky & Cheyann:Modeling Demonstration/Peer Review Draft
- **May 8th:**Ricky & Cheyann: Final Present

Questions?