project_2: e_sports

team_name: pink + pandas.py
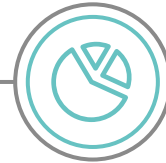
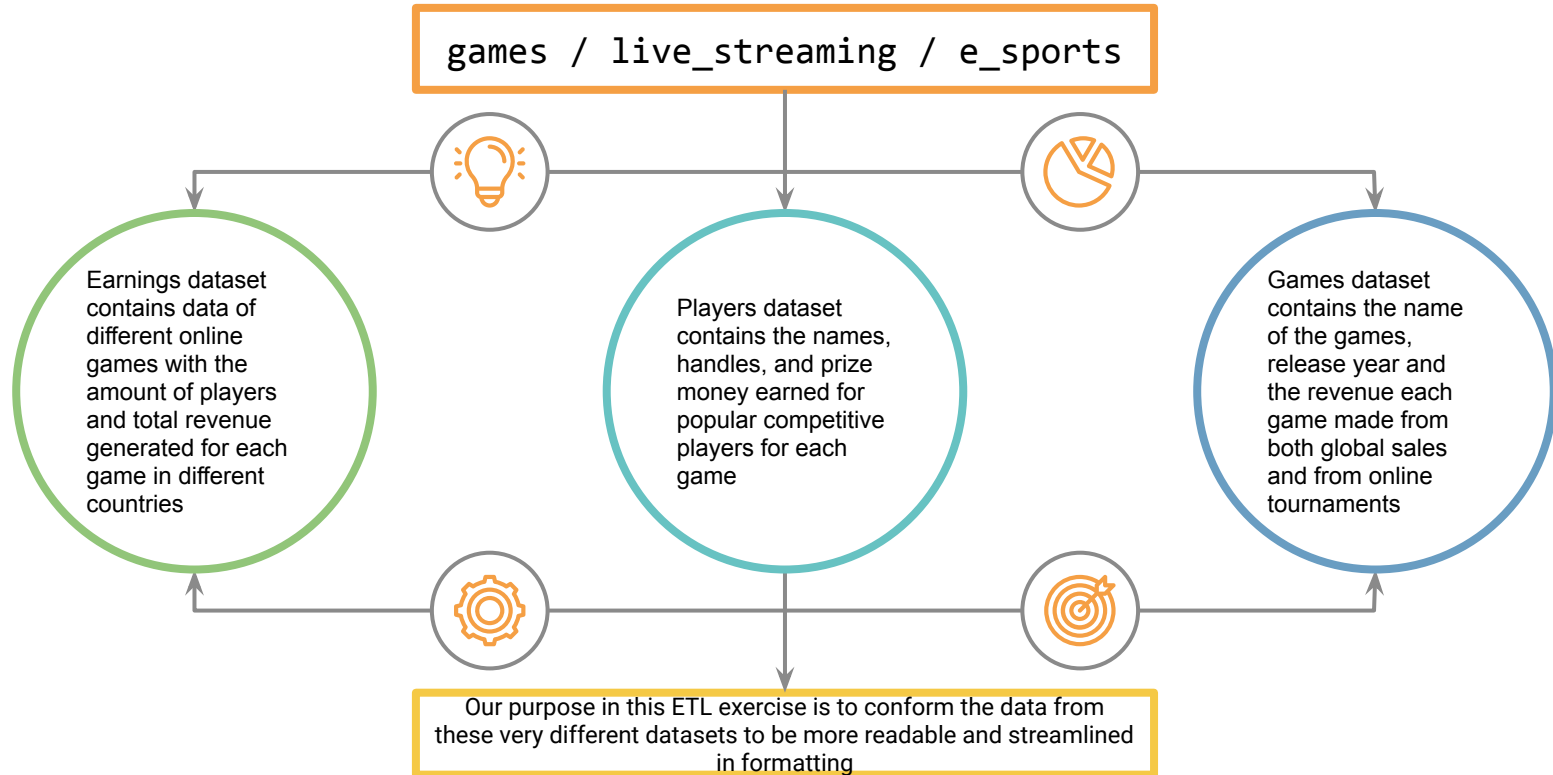<earnings

players
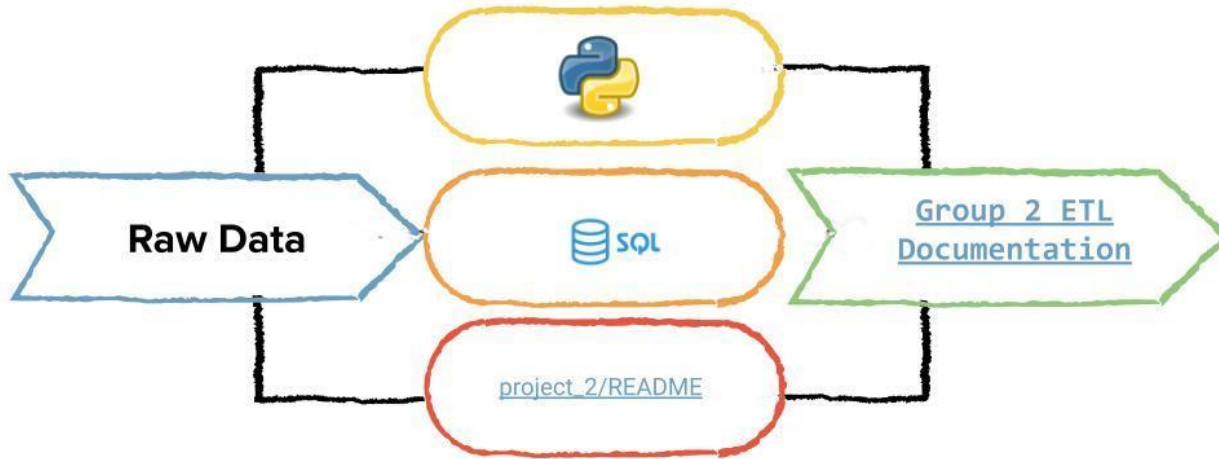
games/>

Project topic: Games   kaggle.com

```
games / live_streaming / e_sports
```

Earnings dataset contains data of different online games with the amount of players and total revenue generated for each game in different countries

Players dataset contains the names, handles, and prize money earned for popular competitive players for each game

Games dataset contains the name of the games, release year and the revenue each game made from both global sales and from online tournaments

Our purpose in this ETL exercise is to conform the data from these very different datasets to be more readable and streamlined in formatting

ETL

Raw Data → [Python] [SQL] [project_2/README] → Group 2 ETL Documentation

[Group 2 ETL Documentation](#)

[project_2/README](#)

IdNo	TotalMoney	GameName	Genre	PlayerNo	TournamentNo	Top_Country	Top_Country_Earnings	Releaseyear
613	310	Acceleration of SUGURI 2	Fighting Game	6	2	United States	310	2018
510	190075.61	Age of Empires	Strategy	246	95	Viet Nam	123408.43	1997
179	1492489.51	Age of Empires II	Strategy	956	576	China	174252.44	1999
177	68113.85	Age of Empires III	Strategy	106	76	United States	18523.52	2005
540	2668	Age of Empires Online	Strategy	16	7	Germany	1264	2011
178	52360	Age of Mythology	Strategy	22	6	Taiwan, Republic of China	22000	2002
667	20	Akatsuki Blitzkampf	Fighting Game	3	1	United States	20	2007
246	40000	Aliens versus Predator 2	First-Person Shooter	1	1	United States	40000	2001
702	11000	Among Us	Strategy	5	2	United States	10000	2018
566	4273351.69	Apex Legends	Battle Royale	1099	170	United States	1825646.05	2019
457	245	Aquapazza: Aquaplus Dream Match	Fighting Game	5	2	United States	128.5	2011
431	1000	Arcana Heart 3	Fighting Game	4	1	United States	950	2009
597	570	Arcana Heart 3: Love Max Six Stars!!!!!	Fighting Game	8	4	United States	464	2014
462	410	Arcana Heart 3: LOVE MAX!!!!!	Fighting Game	6	2	United States	284	2013
529	15228710.62	Arena of Valor	Multiplayer Online Battle Arena	558	51	China	5016882.31	2015
508	18248.09	ARMS	Fighting Game	95	82	United States	11089.37	2017
552	119168.54	Artifact	Collectible Card Game	275	166	China	26273.47	2018
661	155	Artifact 2.0	Collectible Card Game	3	3	Russian Federation	60	2020
521	29500	Asphalt 4	Racing	5	2	Singapore	13000	2008
625	51520	Asphalt 9: Legends	Racing	5	2	Italy	17550	2018
553	191851.7	Assetto Corsa	Racing	81	31	Czech Republic	68689.4	2014
659	1649	Assetto Corsa Competizione	Racing	8	9	Denmark	354	2018

project_2: ETL

Members:
Benji, Meagan, Josh, & Sam

Git Master:
Josh

Topic:
Gaming & E-sports

About:
We like this dataset because it contains a variety of datatypes like strings and integers. We have the option to join our tables on the titles of the games, the genre of the games, or the released date of the games. The datasets provide information on earnings by game and by e-sport player which will provide us clarity on the financial performance of modern-day online games along with their popularity with viewers.

Games:
Minecraft, 2k, Overwatch, League of Legends, R6-S (in general)

Final Production Database:
We chose using PostgreSQL over a NoSQL Database because our tables will require a lot of joining and a defined standard schema so we feel more comfortable with using a relational Database.

Datasets:
Esports Players Earnings csv:
https://www.kaggle.com/jackdaoud/esports-earnings-for-players-teams-by-game?select=highest_earning_players.csv

Esports Earnings 1998-2021:
https://www.kaggle.com/rankirsh/esports-earnings
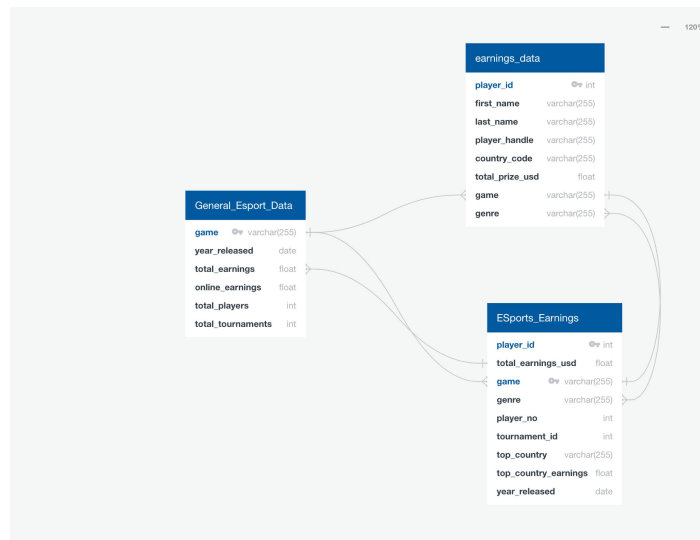
Business Rules:

--- table level rules ----

- Keep gaming data related to the US, filter out other countries
- Only keep data for players that only over $700k
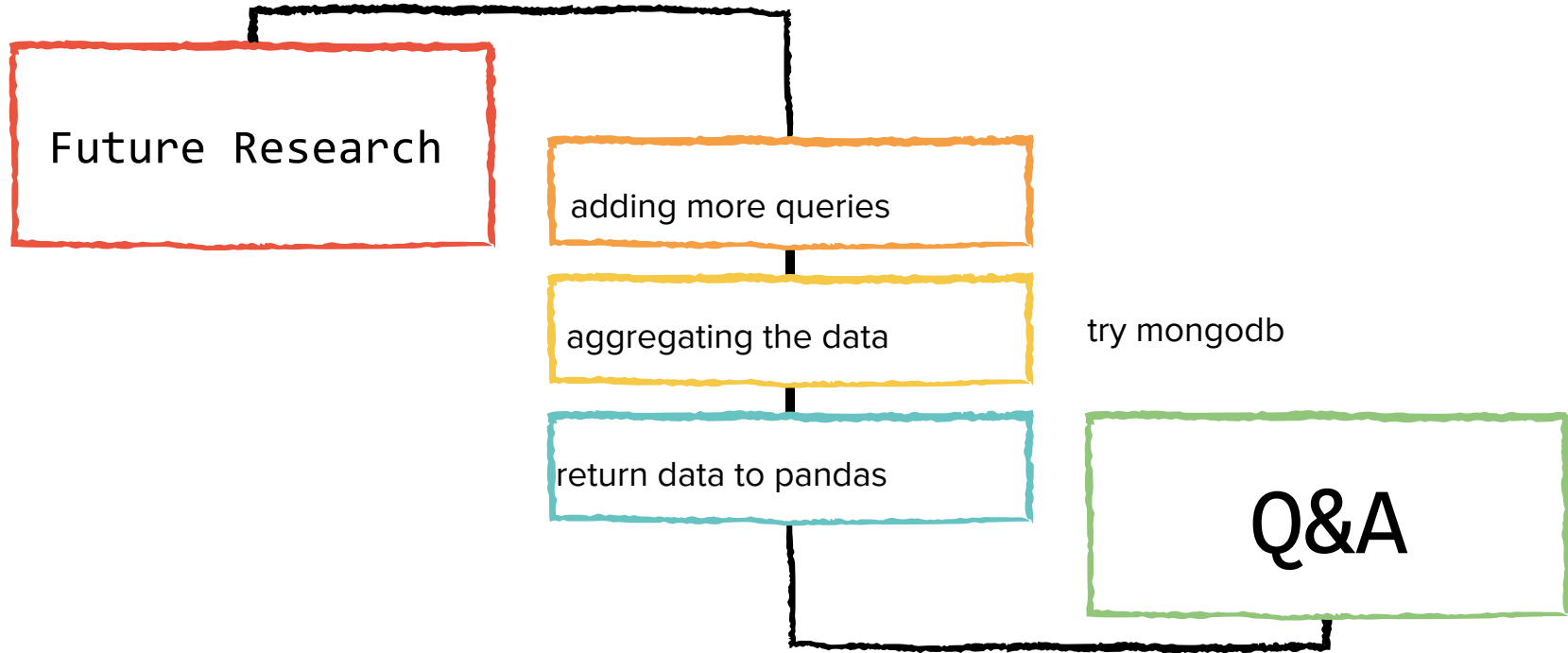- drop any row that has any missing data, only complete data

---- column level rules ----

- lower case
- format date
- unformat numbers --> INT

Organizing Visuals for Presentation:
https://slidesgo.com/theme/linear-flowchart-infographics

Future Research

adding more queries

aggregating the data

return data to pandas

try mongodb

Q&A

<Thank you!/>