LOADING...

Please remember
NDA slide from Chris!

Project X

EARLY GAME DEVELOPMENT DATA ANALYSIS

Life Beyond developed by Darewise Entertainment

(alpha version released in late 2019)

Scenario:

- Darewise is a young gamedev studio.
- Life Beyond recently entered alpha testing and lacks data analysis. CEO wants to have as much insights from what they have as possible.
- Added value to game designers and everyone involved.

Objectives:

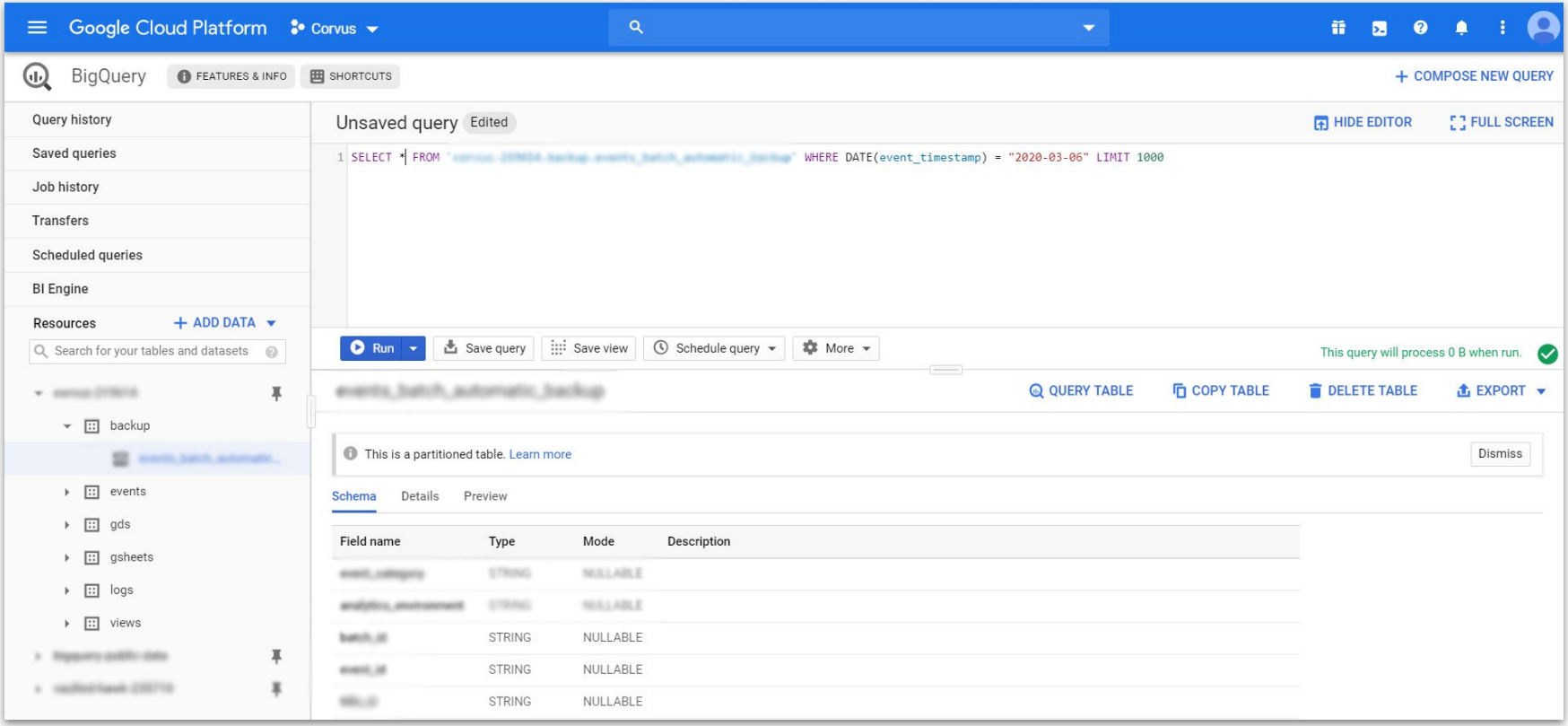
Get as much data as possible

Explore and analyze everything in Life Beyond

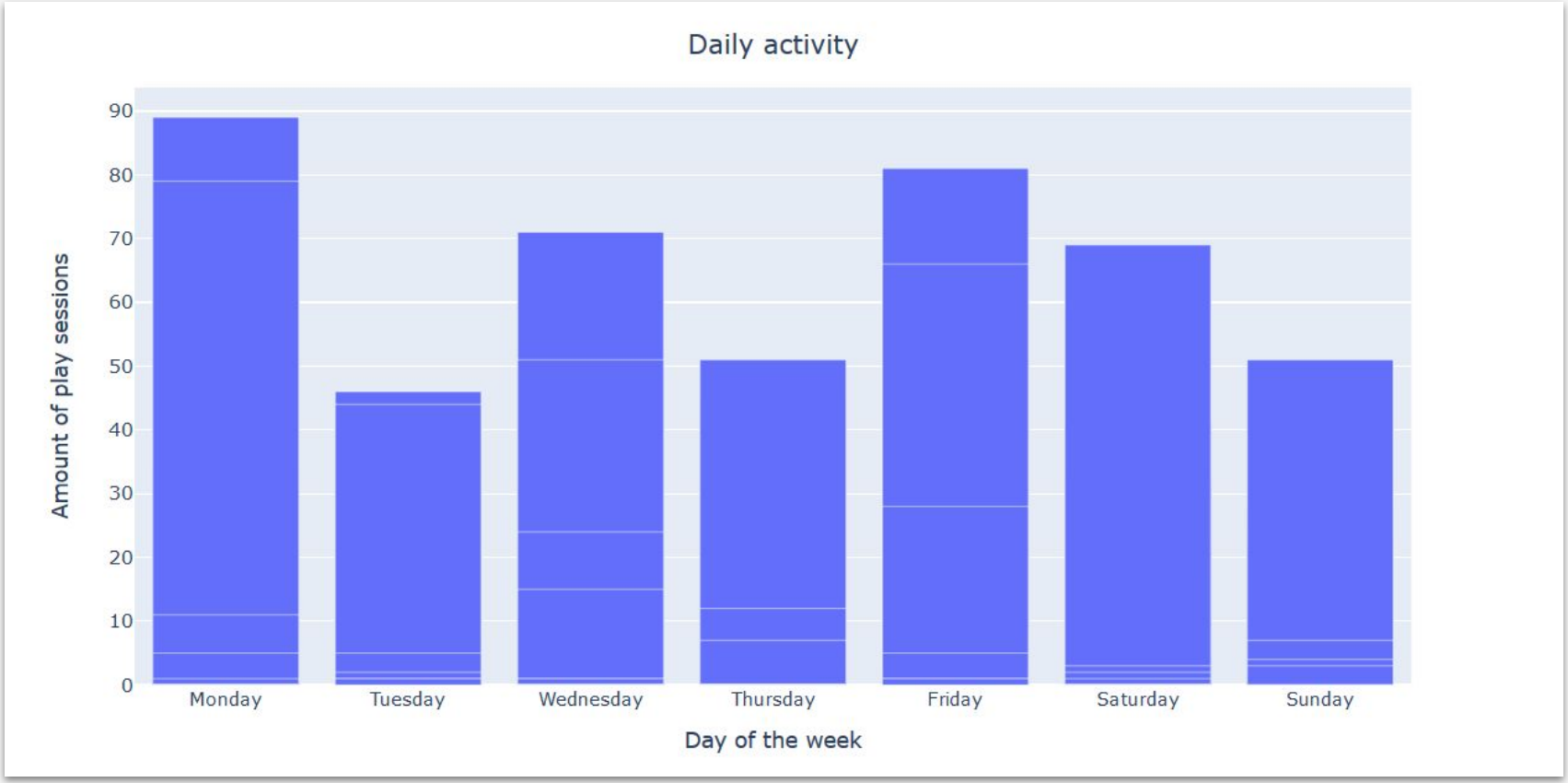
Visualize insights and useful findings

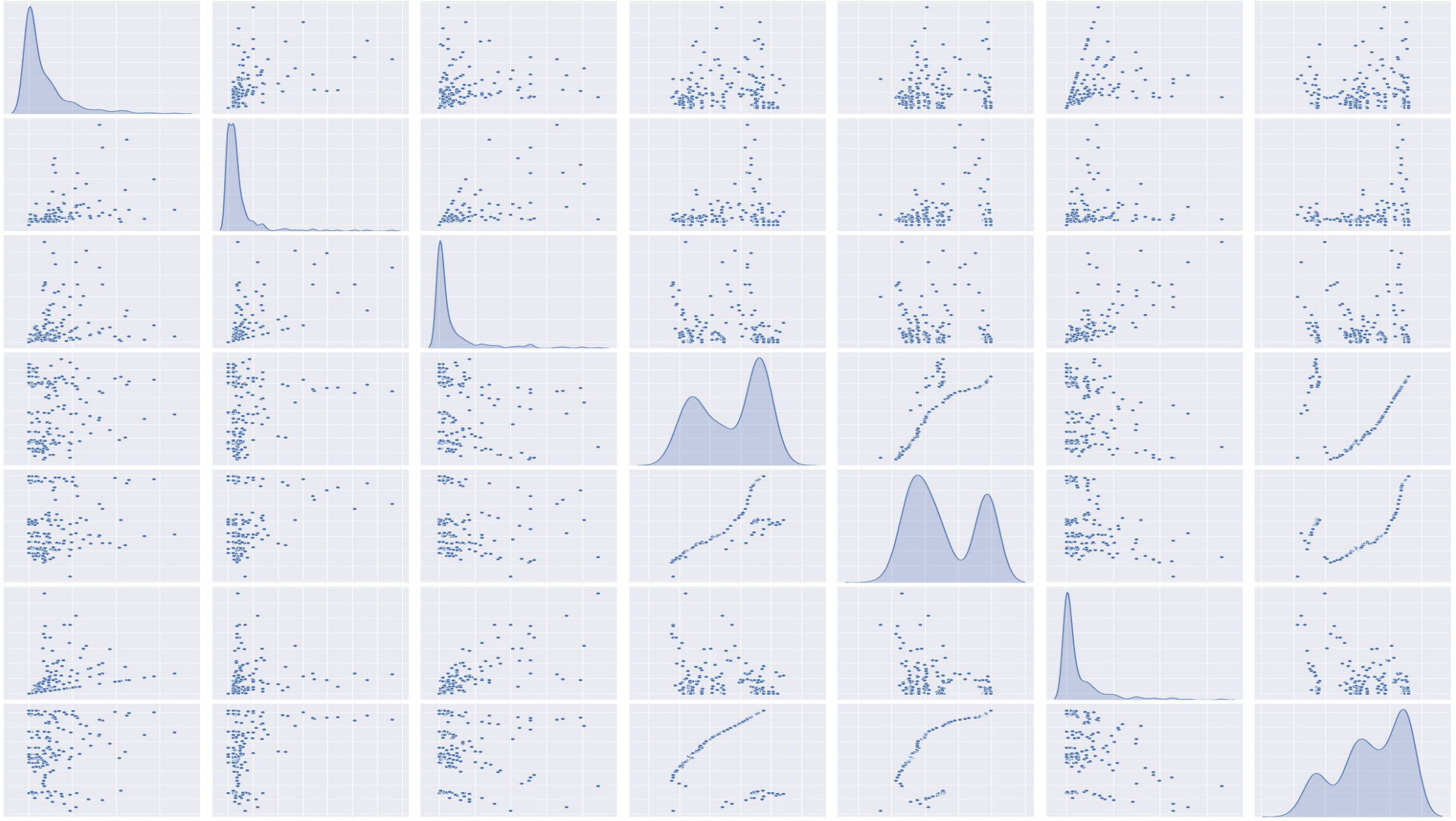
Data collection

- Data generated by game engine, processed by Microsoft Azure PlayFab, stored in Google Cloud within BigQuery
- Python connection to BigQuery
- Unwrapping and cleaning

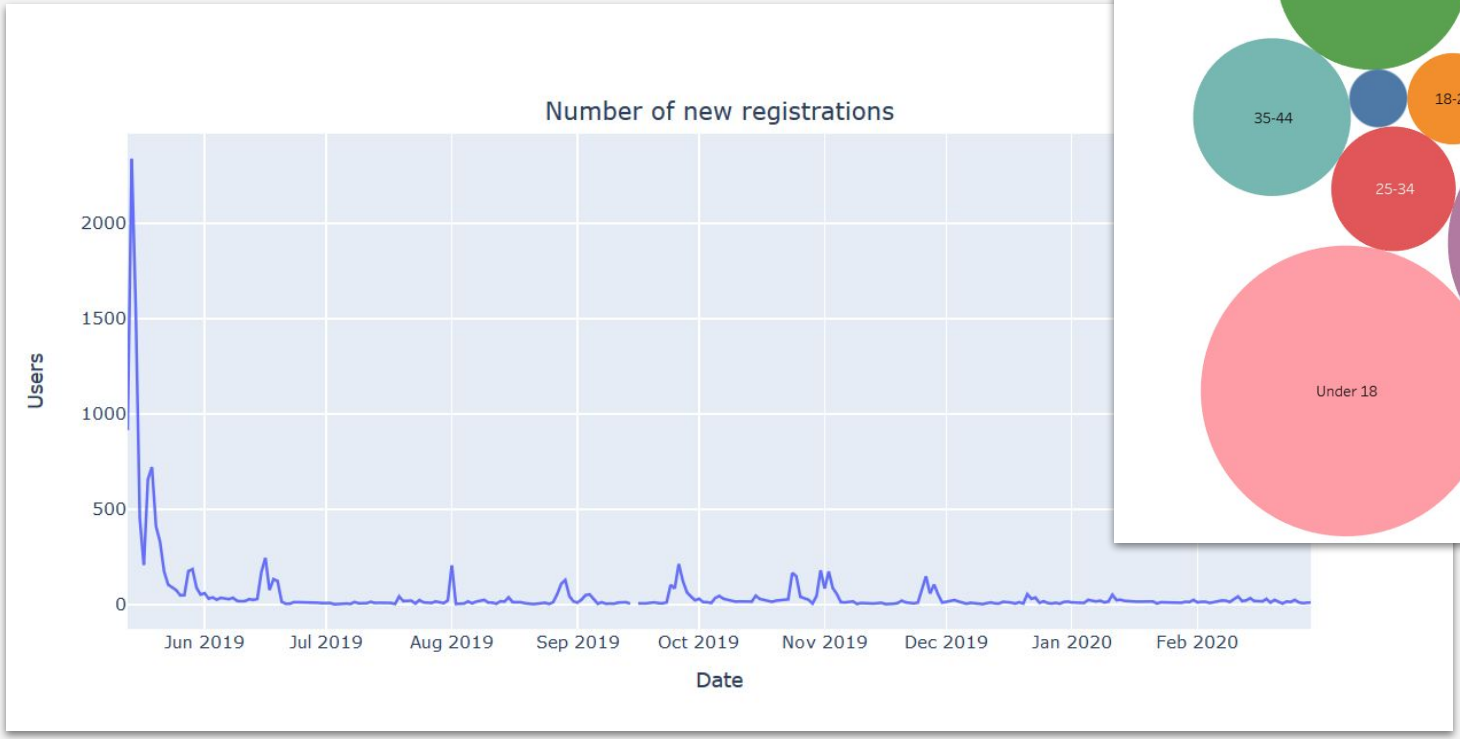


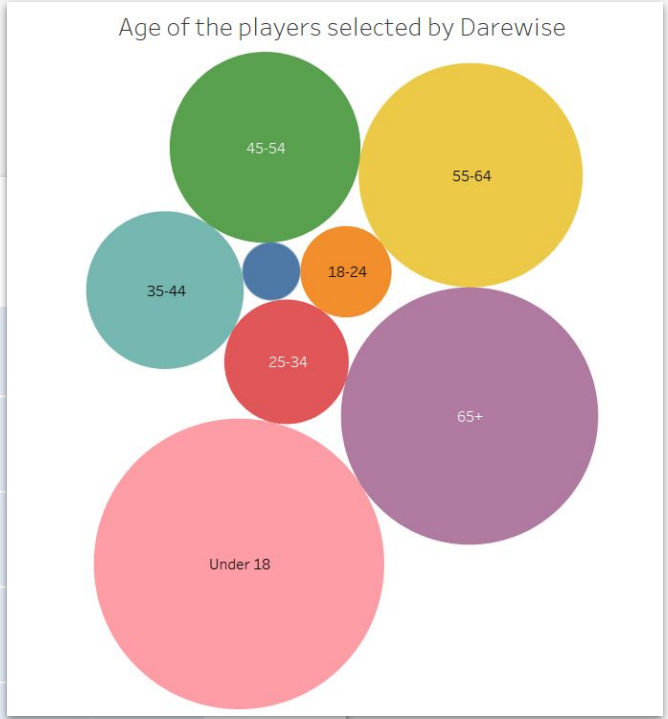
Data visualization



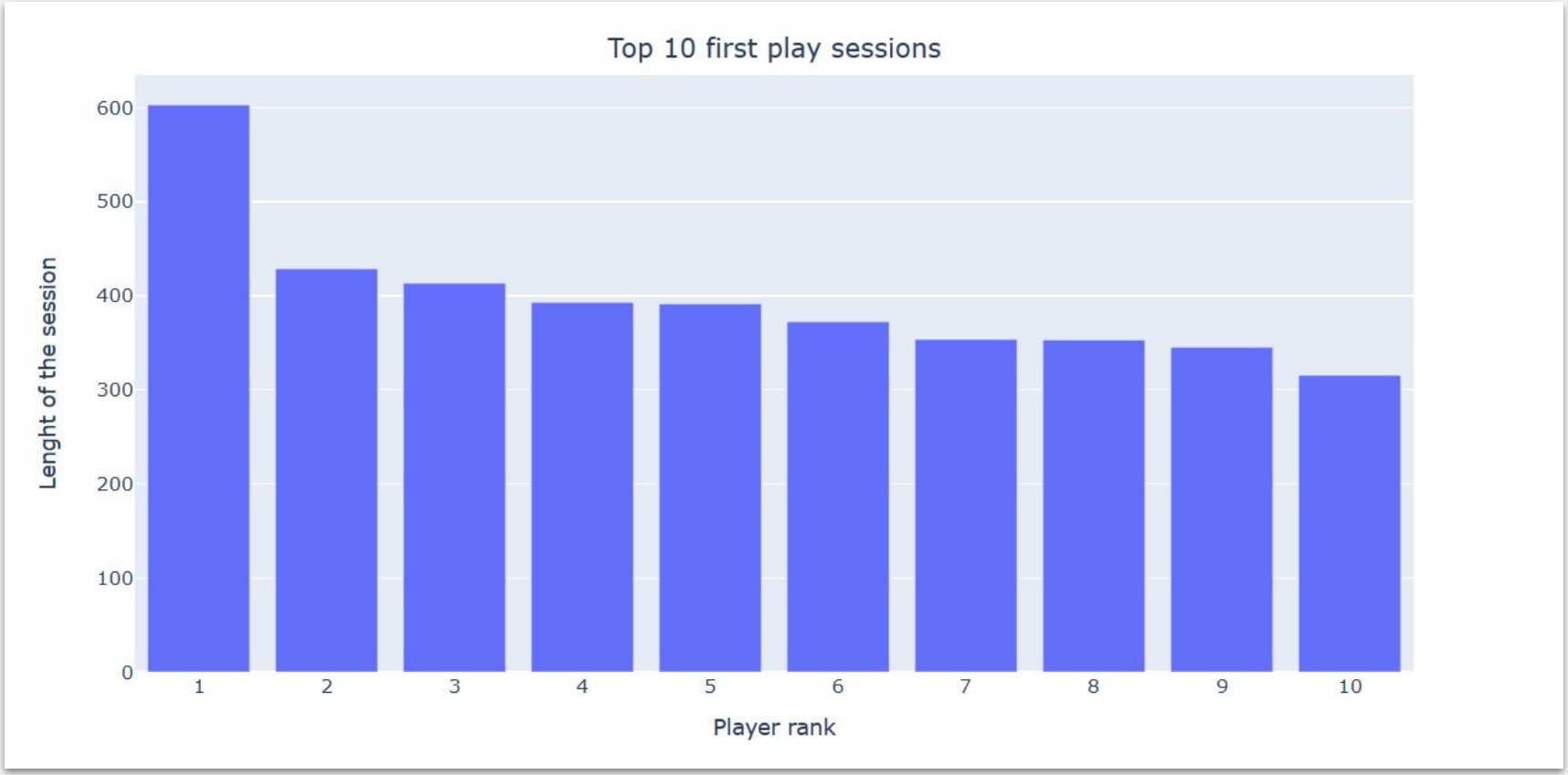


Data visualization

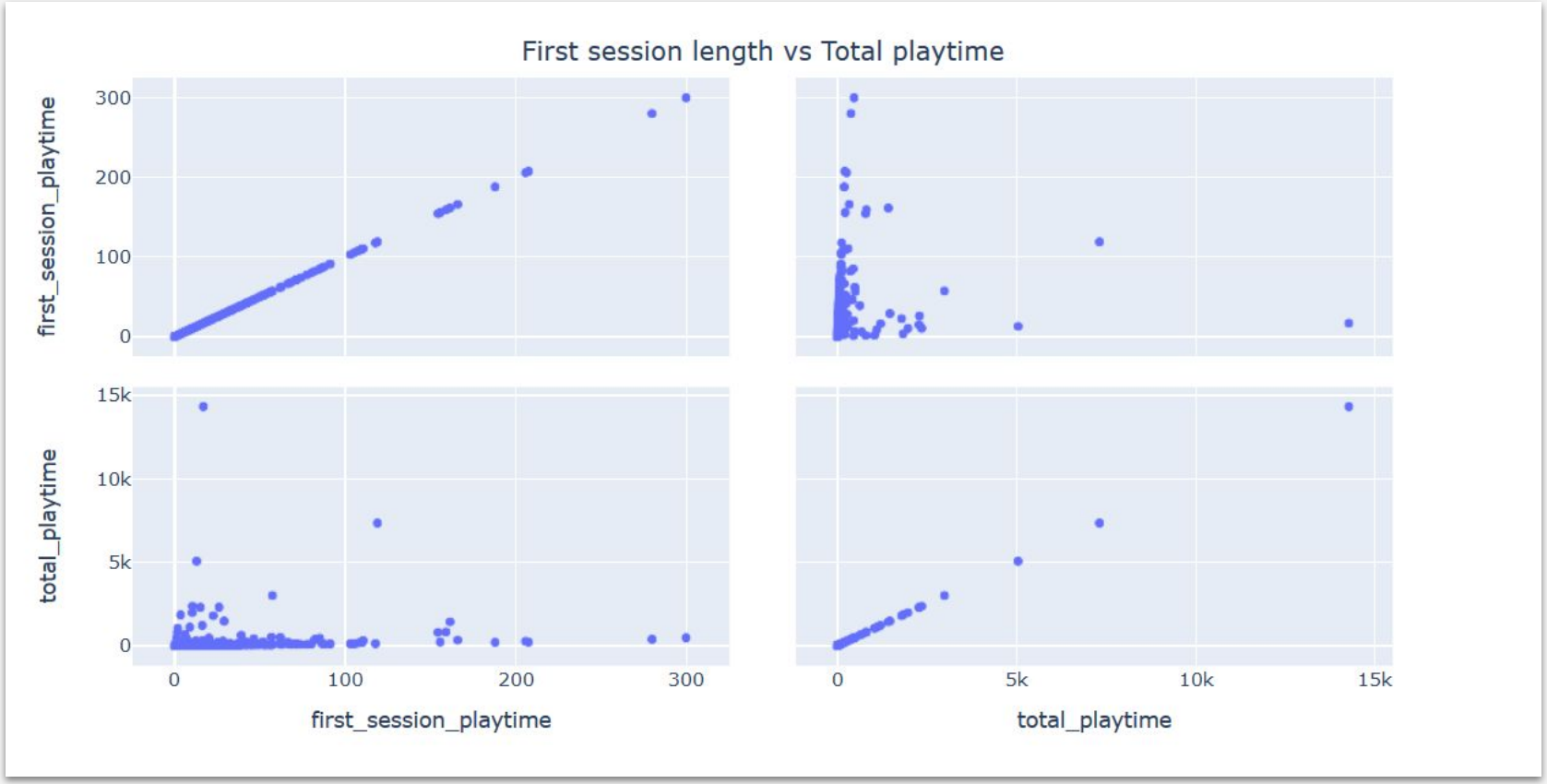




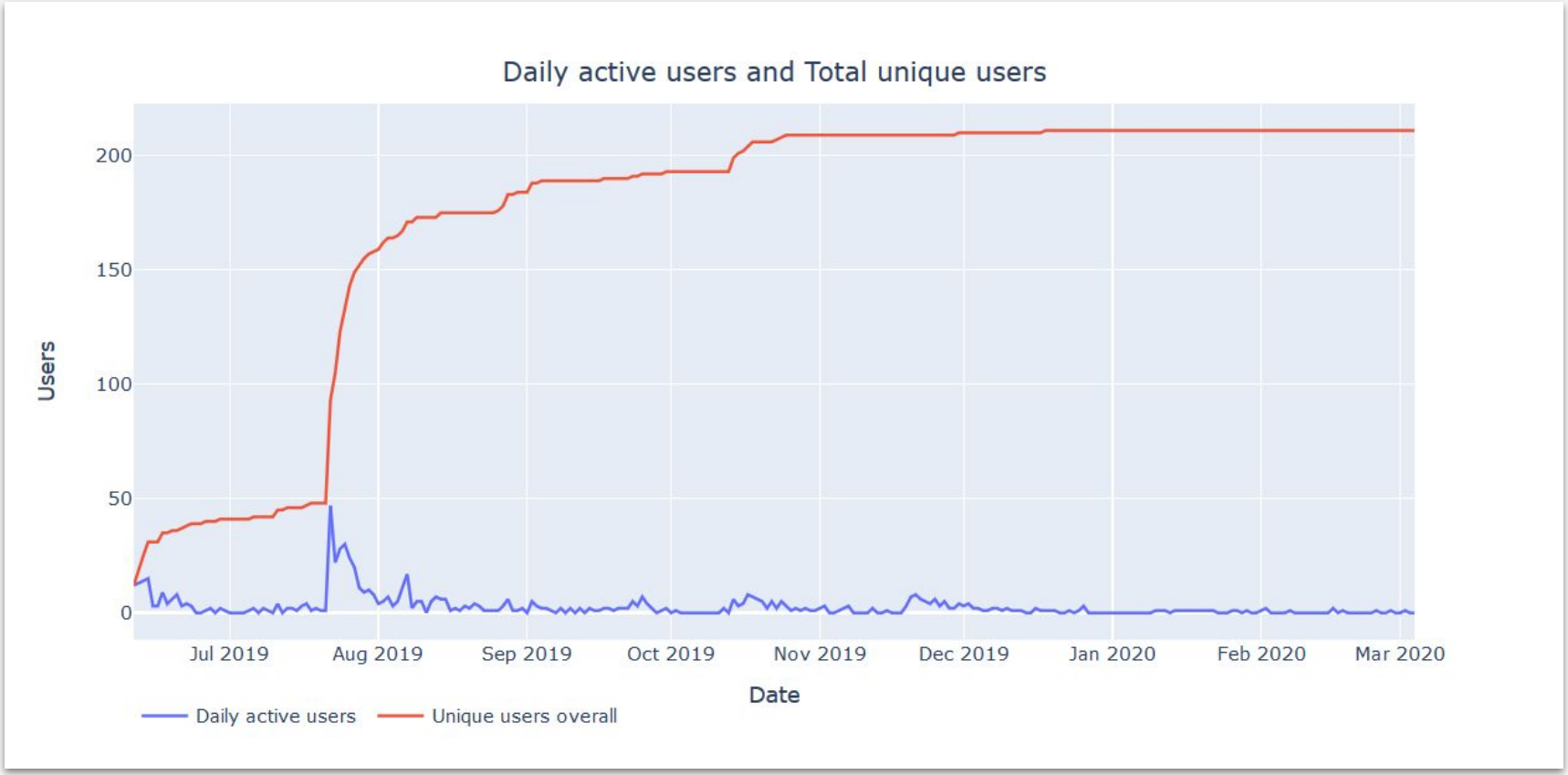
Data visualization

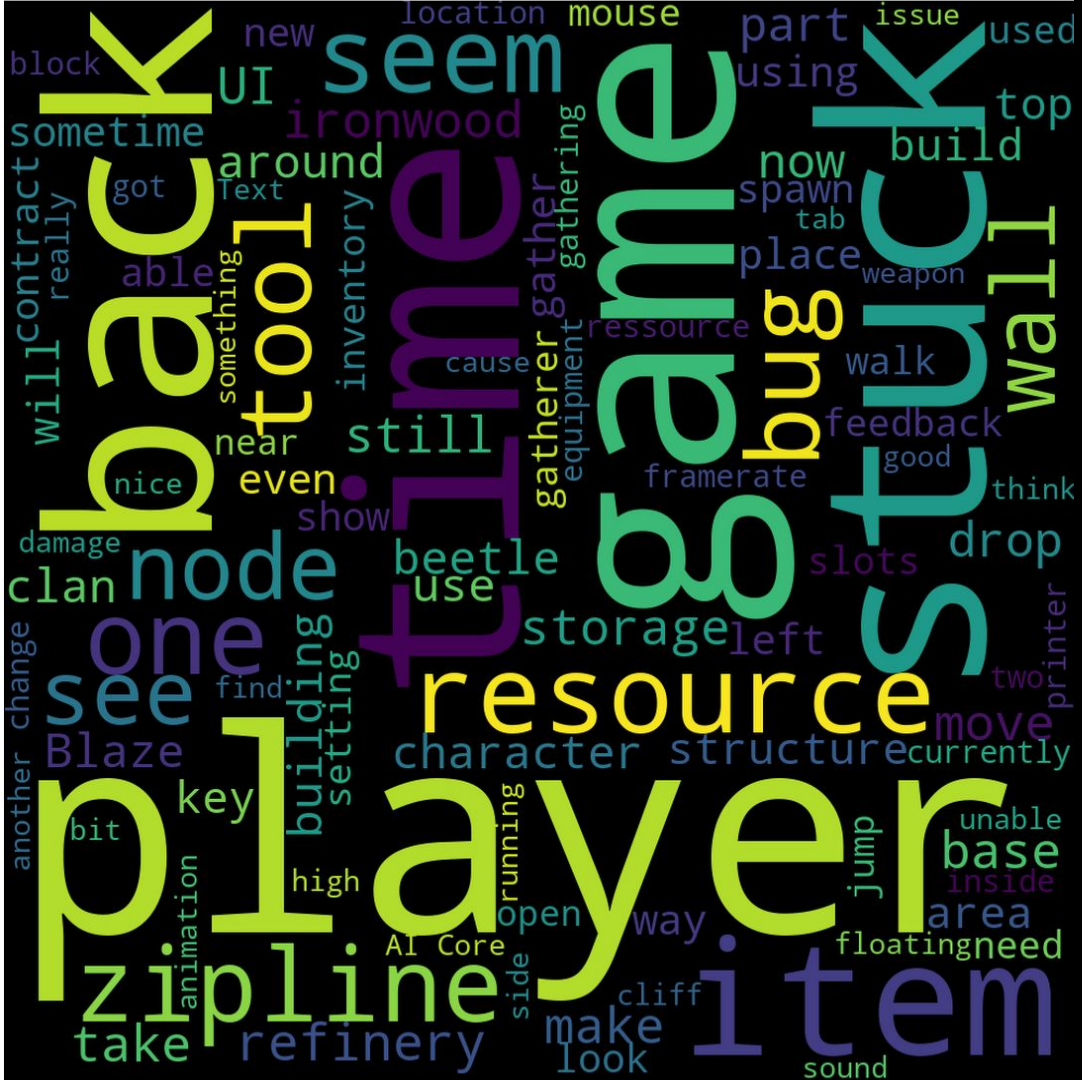


Data visualization



Data visualization





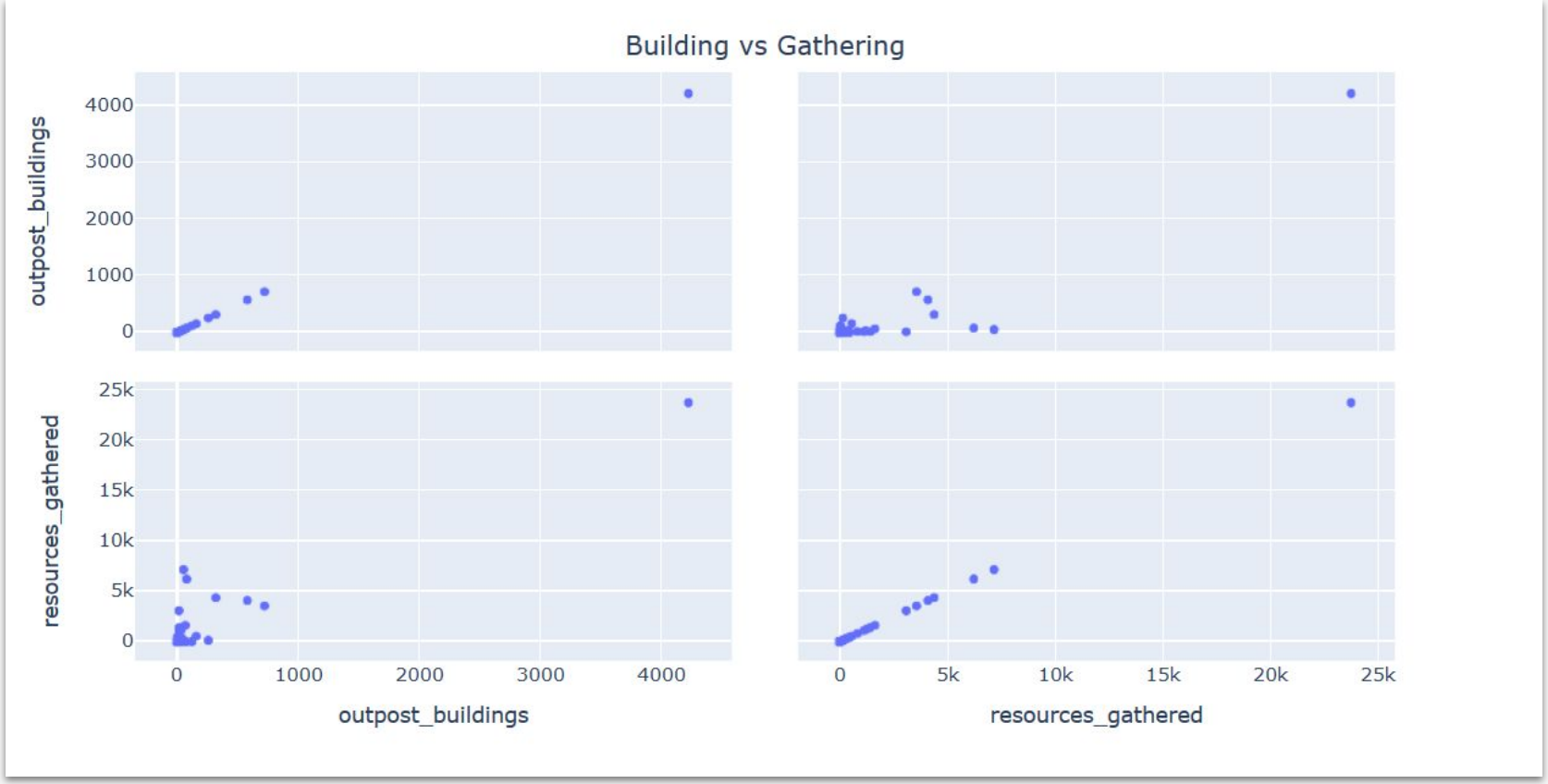
Top 100 words
used by players
in their feedback

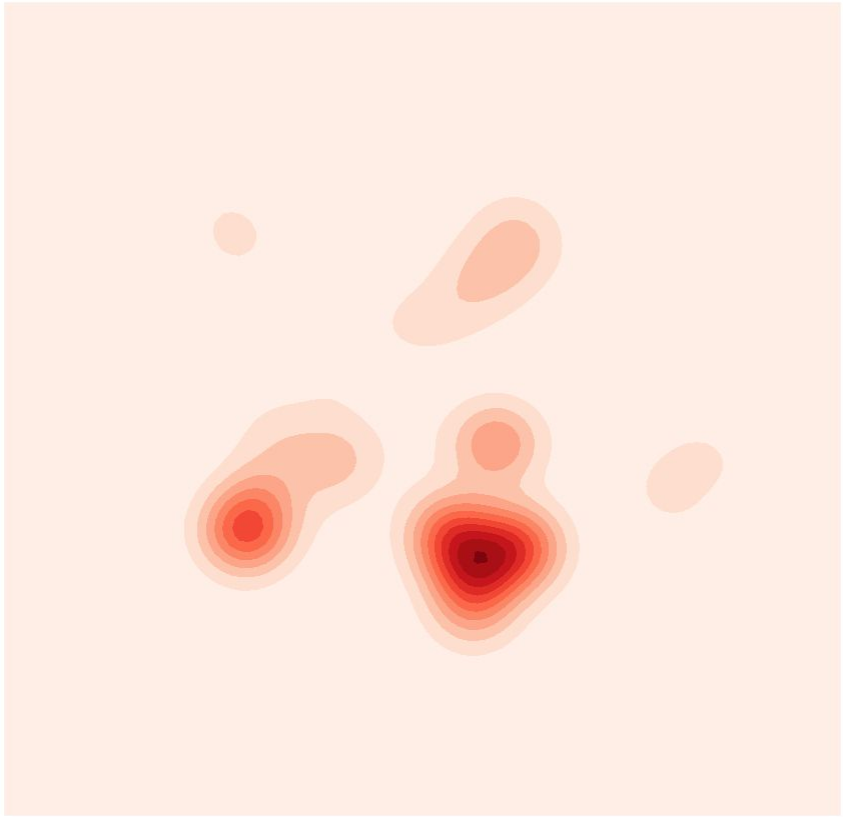
Data cleaning

- JSON normalization hell
- Lack of meaningful data from inside the game
- Missing data descriptors

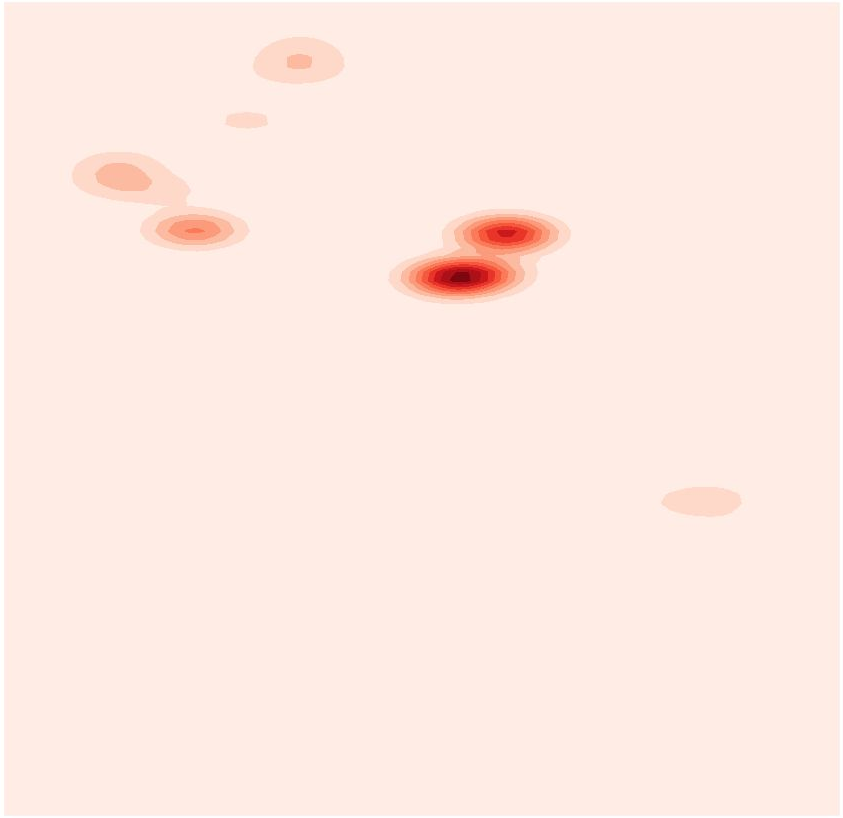
	Influencer	Wave	Status	Age	Platform	Agree_to_help	Community_member	Grand Theft Auto Online	Assassin's Creed Odyssey	World of Warcraft	ARK: Survival Evolved	Currently_playing	Playing_with_friends
4883	0	5.0	1	2.0	3.0	1	3	0.0	0.0	0.0	0.0	Tom Clancy's Ghost Recon Wildlands. Playing it...	1.0
2596	0	0.0	0	2.0	3.0	1	3	1.0	1.0	1.0	1.0	Currently I Am Playing Guild Wars 2 because i ...	3.0
5182	0	0.0	0	2.0	2.0	1	3	1.0	1.0	0.0	1.0	Skyrim, because Monday is for modding.	1.0
541	0	0.0	0	2.0	2.0	1	1	1.0	0.0	1.0	0.0	Disgaea 5 due to how much content there is to ...	1.0
1132	0	0.0	0	3.0	3.0	1	1	1.0	1.0	1.0	1.0	there are 2 games that I play currently play d...	2.0
7097	0	4.0	1	4.0	3.0	1	2	1.0	1.0	0.0	1.0	Ark, Assassins Creed Odyssey, Conan Exiles, Th...	3.0
4369	0	0.0	0	3.0	1.0	1	1	0.0	1.0	1.0	0.0	Blade and soul, because im bored of all other ...	2.0
7055	0	0.0	0	2.0	3.0	1	3	1.0	0.0	0.0	0.0	Rainbow Six Siege	2.0
2665	0	0.0	0	4.0	3.0	1	3	1.0	1.0	1.0	1.0	Beamng.drive, i love physics.	0.0

More data visualization

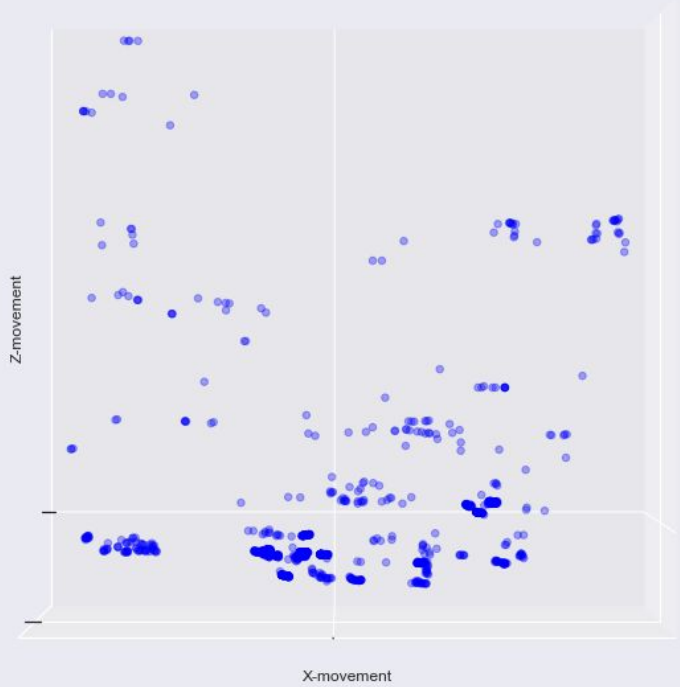




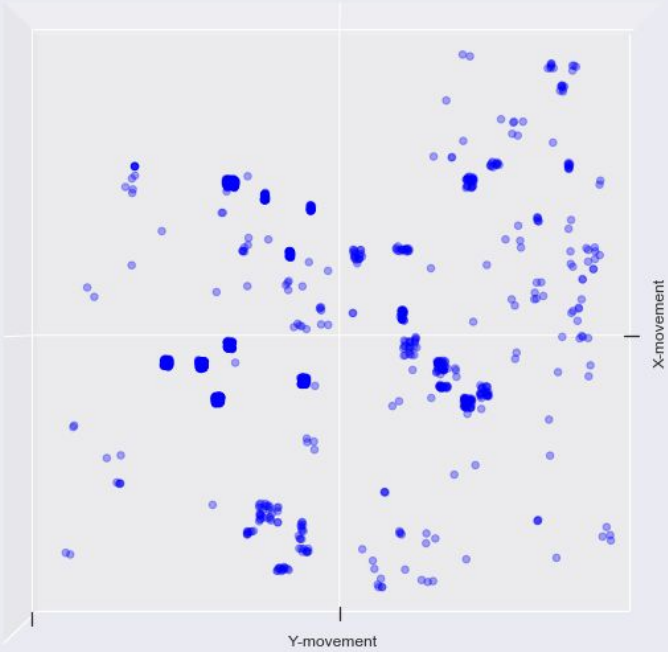
Heatmap
(players actions
defined by XY coords
in coordinates set_1)



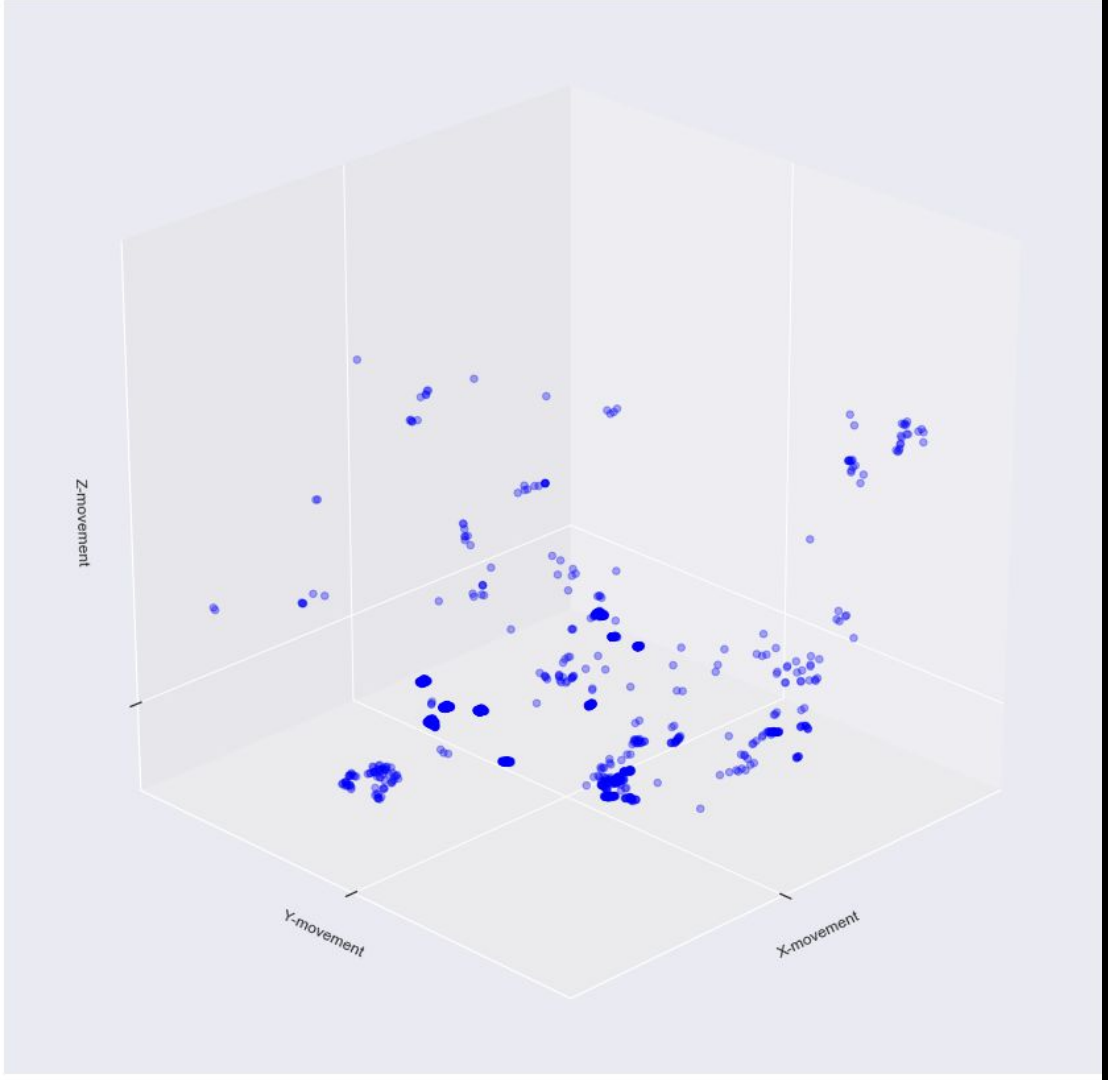
Heatmap #2
(players actions
defined by XY coords
in coordinates set_2)



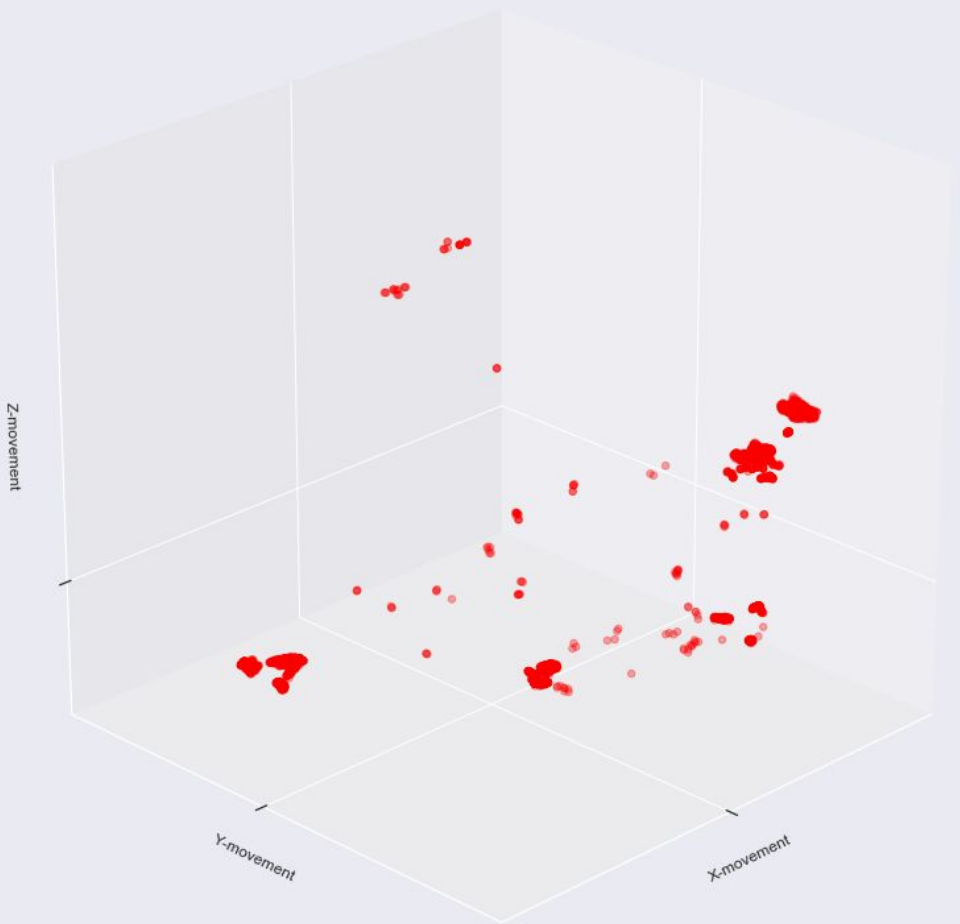
3D plot of players actions
XZ axes
coordinates set_1



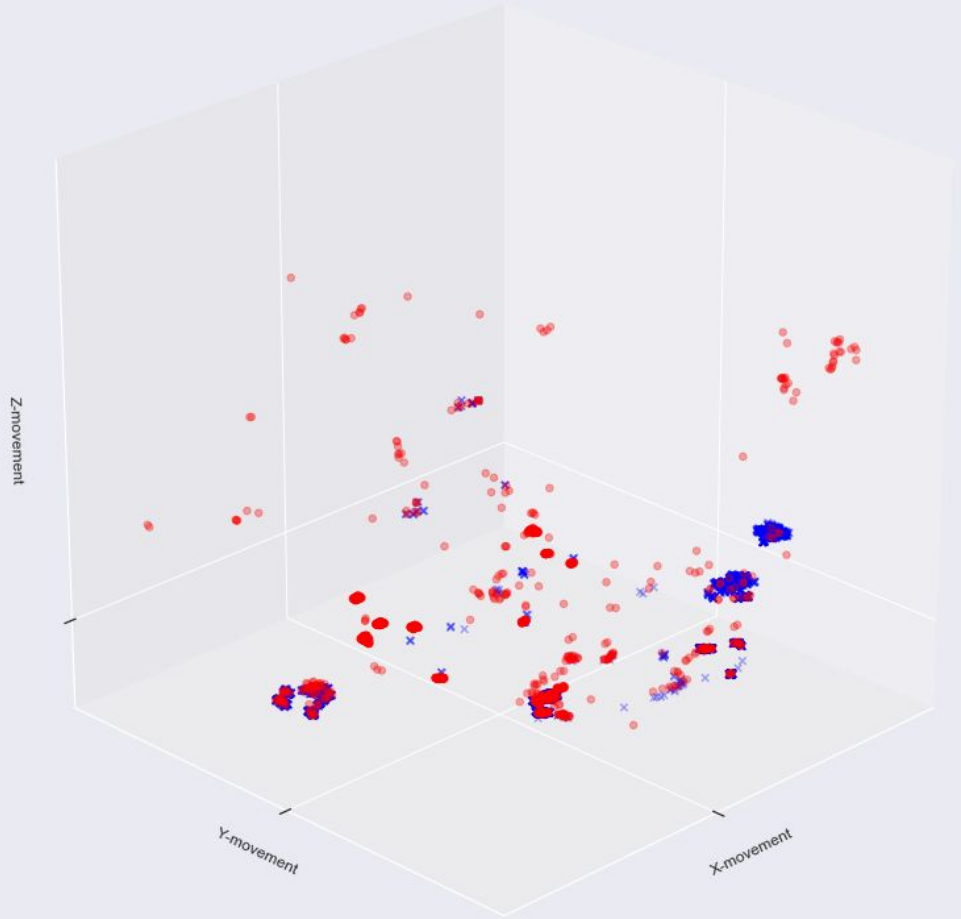
3D plot of players actions
XY axes
coordinates set_1



3D plot of players actions
XYZ axes
coordinates set_1



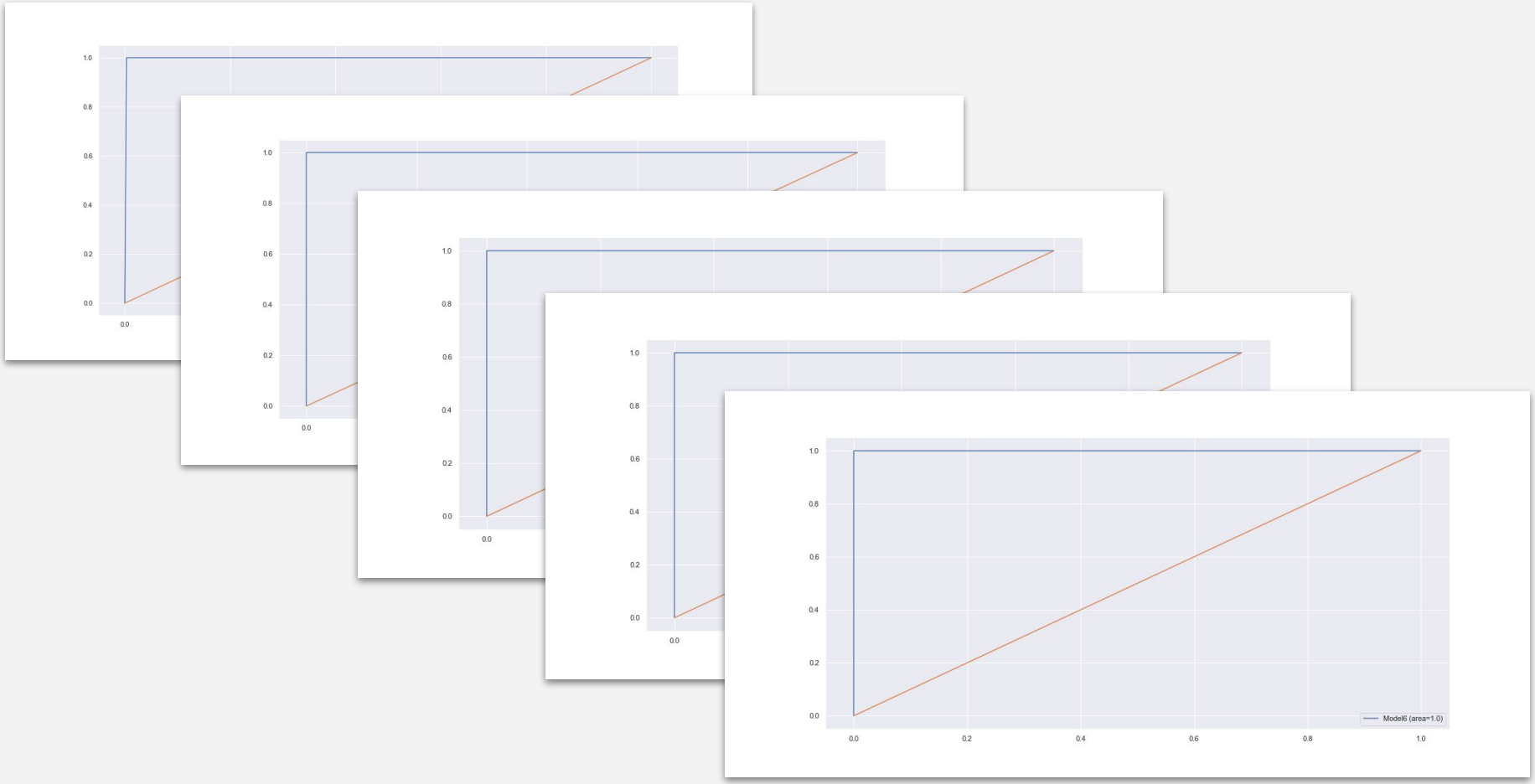
3D plot of players actions
XYZ axes
coordinates set_2



3D plot of players actions
XYZ axes
two coordinate sets compared

Models

Models - predict if player is going to provide feedback



Accuracy stats overview

	Accuracy	F1	AUC	Recall	Precision
Naive Bayes	0.9996	0.9998	0.9984	1.0000	0.9996
LinearSVC	1.0000	1.0000	1.0000	1.0000	1.0000
KNeighbors	0.9996	0.9998	0.9984	1.0000	0.9996
LogisticRegression	1.0000	1.0000	1.0000	1.0000	1.0000
DecisionTree	1.0000	1.0000	1.0000	1.0000	1.0000
RandomForest	1.0000	1.0000	1.0000	1.0000	1.0000
SVC	0.9996	0.9998	0.9984	1.0000	0.9996
NuSVC	0.9996	0.9998	0.9984	1.0000	0.9996

Summary and lessons learned

- Kaggle is a “perfect scenario” for getting data. Real life cases are not.
- Data that will be recorded needs some ideas put into it beforehand.
- Darewise would be *wise* to go ahead with expansion of its data recording force. Insights come from a meaningful data, and the sooner collection of such things is implemented, the better developers would be able to understand their players.
- Explore the possibilities of fixing deeply nested data from UnrealEngine
- Find more data to make Supervised ML more efficient
- Unsupervised ML
- Move on sooner