HOW TO BUILD A SUCCESSFUL KICKSTARTER CAMPAIGN

Tristan Dresbach

NYC Data Science Academy

10/18/2018

WEB SCRAPING PROCESS
A) SETTING UP SCRAPY

WEB SCRAPING PROCESS
B) CHALLENGES

ISSUES:

1) JavaScript Element

2) 3 Types of Projects

3) IP Ban

SOLUTIONS:

1) ScrapySplash

2) Testing all XPaths

3) Ran on alternate machine

DATA OVERVIEW
A) DATA FRAME

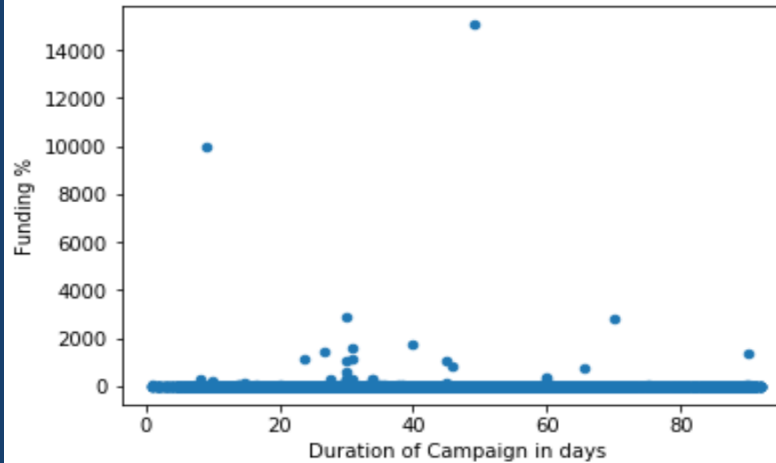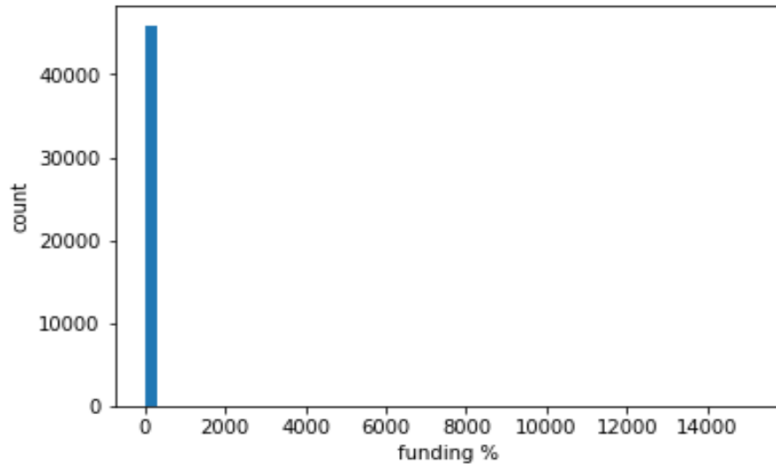|	category	subcategory	location	status	goal	pledged	backers	levels	reward levels	updates	comments	duration
0	Art	Illustration	Chicago, IL	successful	2500.00	2767.0	30	8	1,5,25,50,100,250,500,1,000	7	0	76.96
1	Design	Design	Chicago, IL	successful	7500.00	11331.0	245	8	1,10,20,30,40,50,300,600	12	87	31.96
2	Publishing	Publishing	Pittsburgh, PA	failed	1500.00	85.0	4	4	1,25,50,100	0	0	31.96
3	Music	Music	Sioux Falls, SD	successful	4000.00	7625.0	83	14	8,20,30,40,50,75,100,200,300,400,500,1,000,1,5...	2	1	27.92
4	Theater	Theater	Goshen, IN	successful	5000.00	5010.0	18	13	1,5,10,15,20,40,50,100,150,250,500,1,000,5,000	3	2	31.96
5	Film & Video	Short Film	Baton Rouge, LA	failed	5000.00	60.0	4	6	5,10,25,100,500,2,500	0	0	91.96
6	Film & Video	Film & Video	Dallas, TX	failed	5000.00	5.0	1	7	25,50,100,250,500,750,1,000	6	0	31.96
7	Music	Pop	Brooklyn, NY	failed	10000.00	2055.0	31	6	1,15,50,150,300,1,000	0	1	91.96

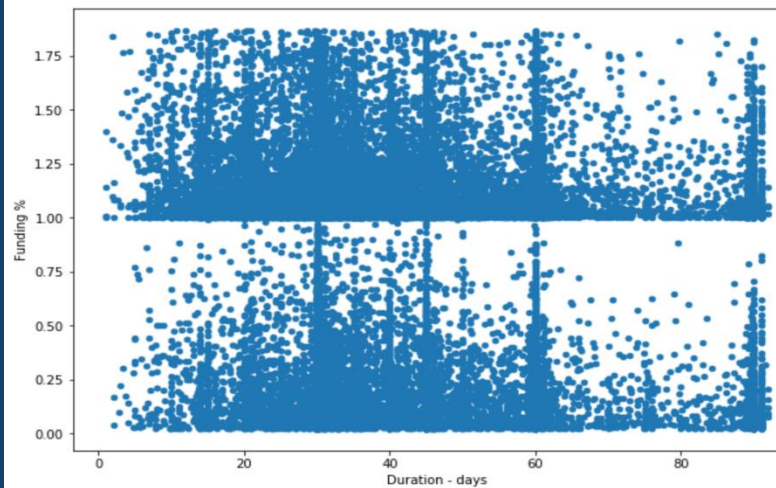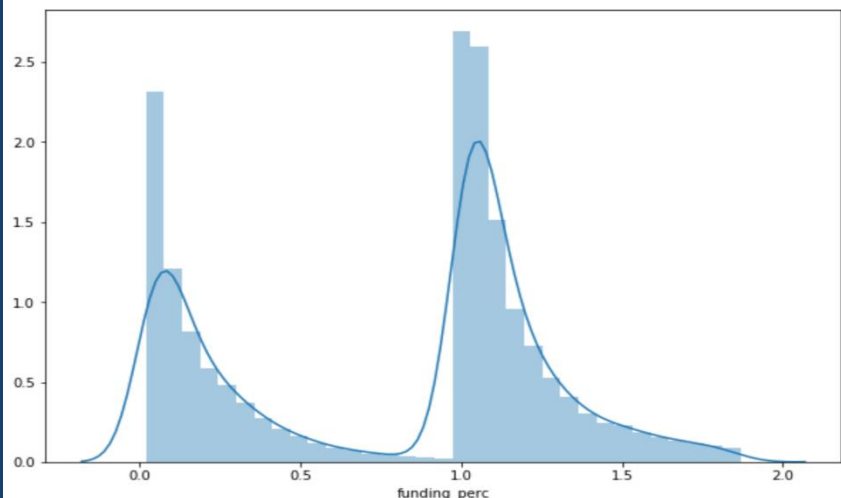▪ What we want to explore: success = (amount funded / goal amount) * 100

DATA OVERVIEW
B) EXAMINING SUCCESS RATES

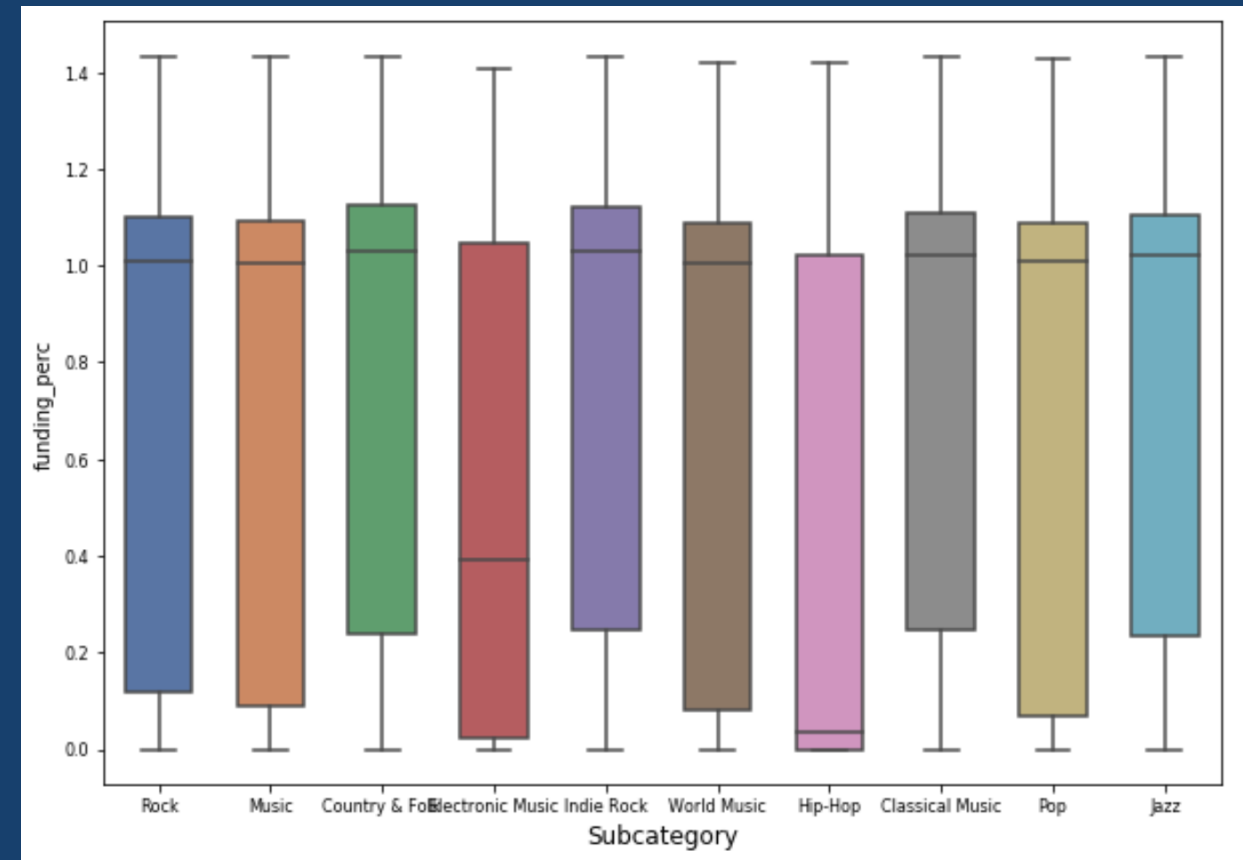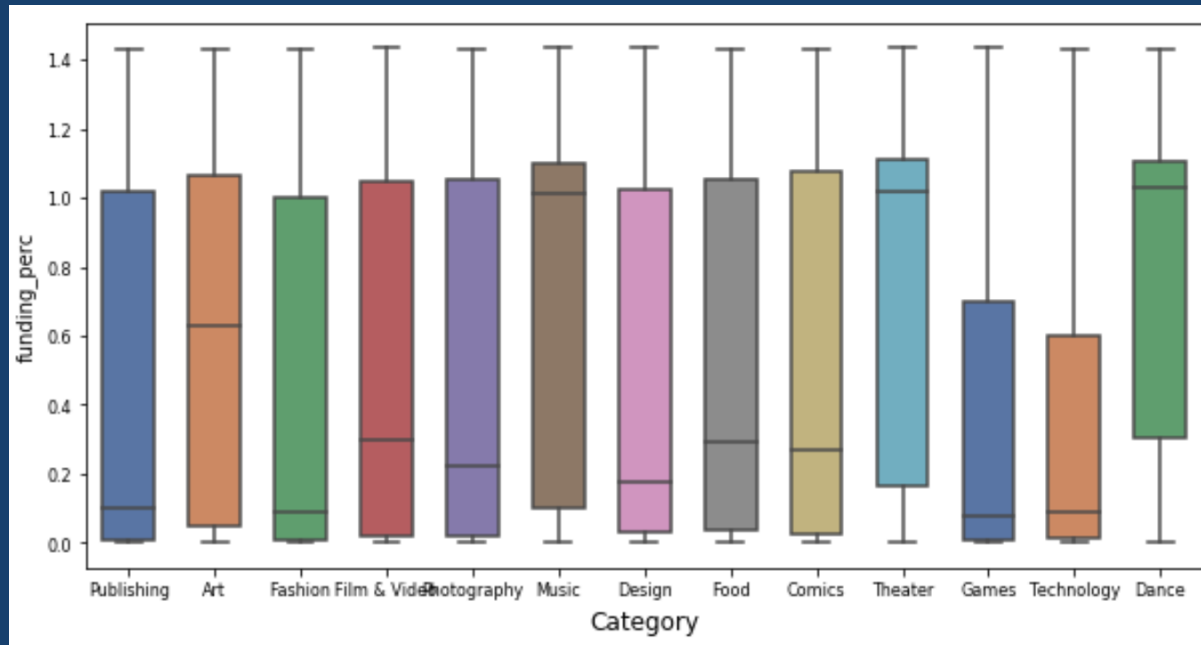Need to do an IQR transformation to get rid of outliers
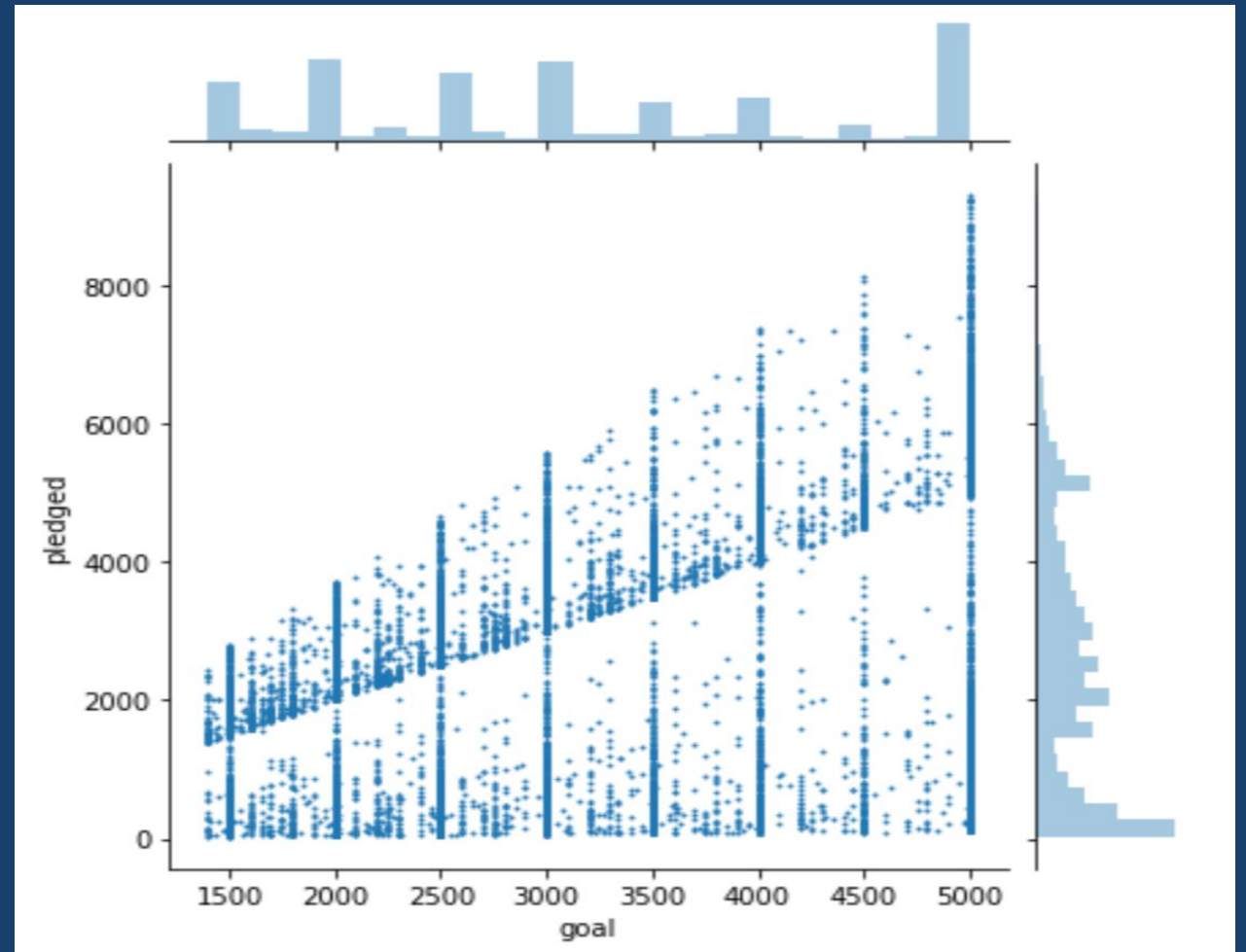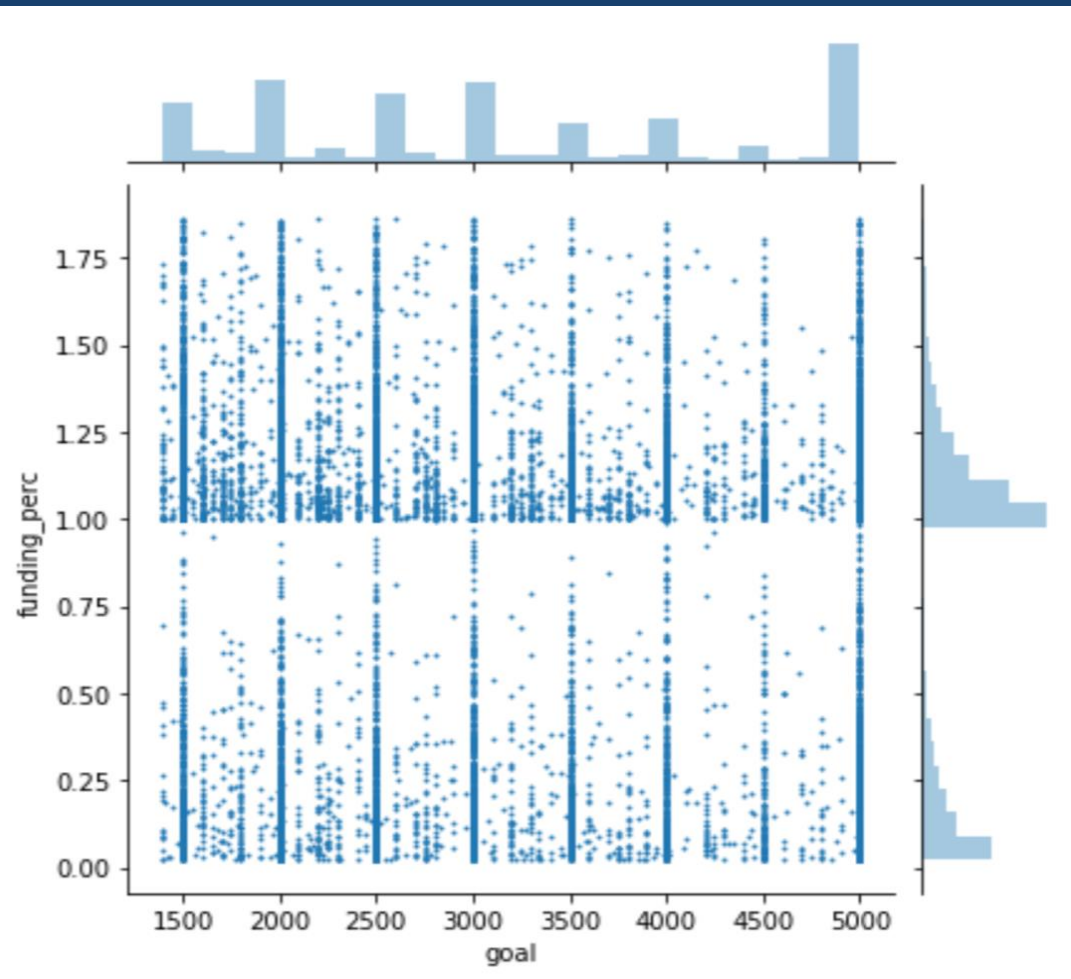
Before IQR
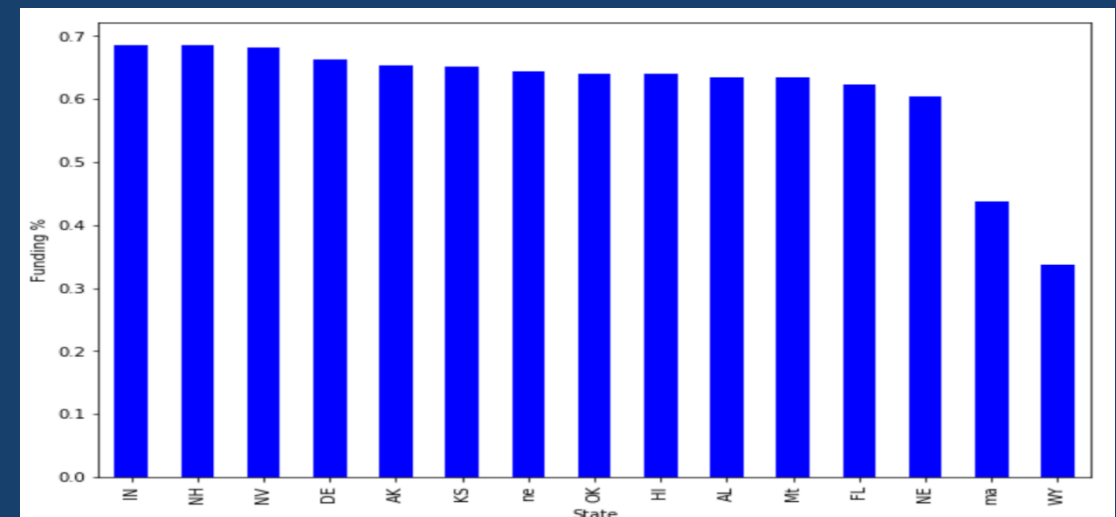
After IQR

DATA ANALYSIS
A) MAIN PROJECT CHARACTERISTICS
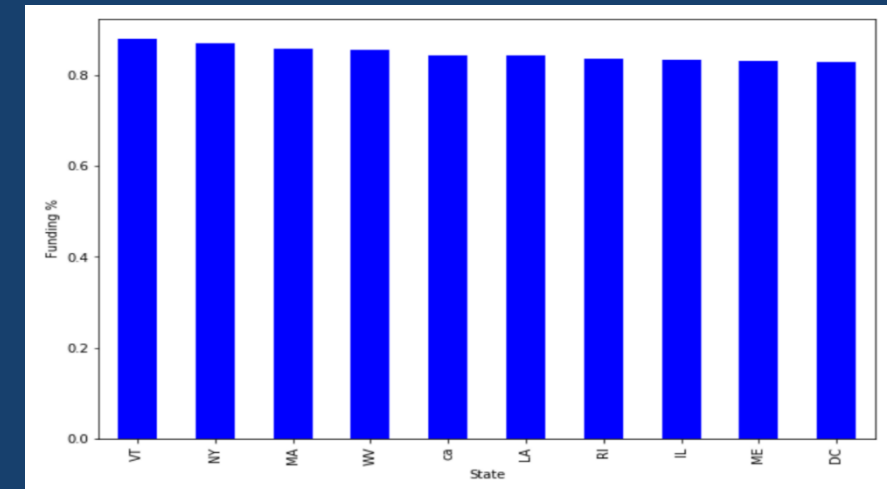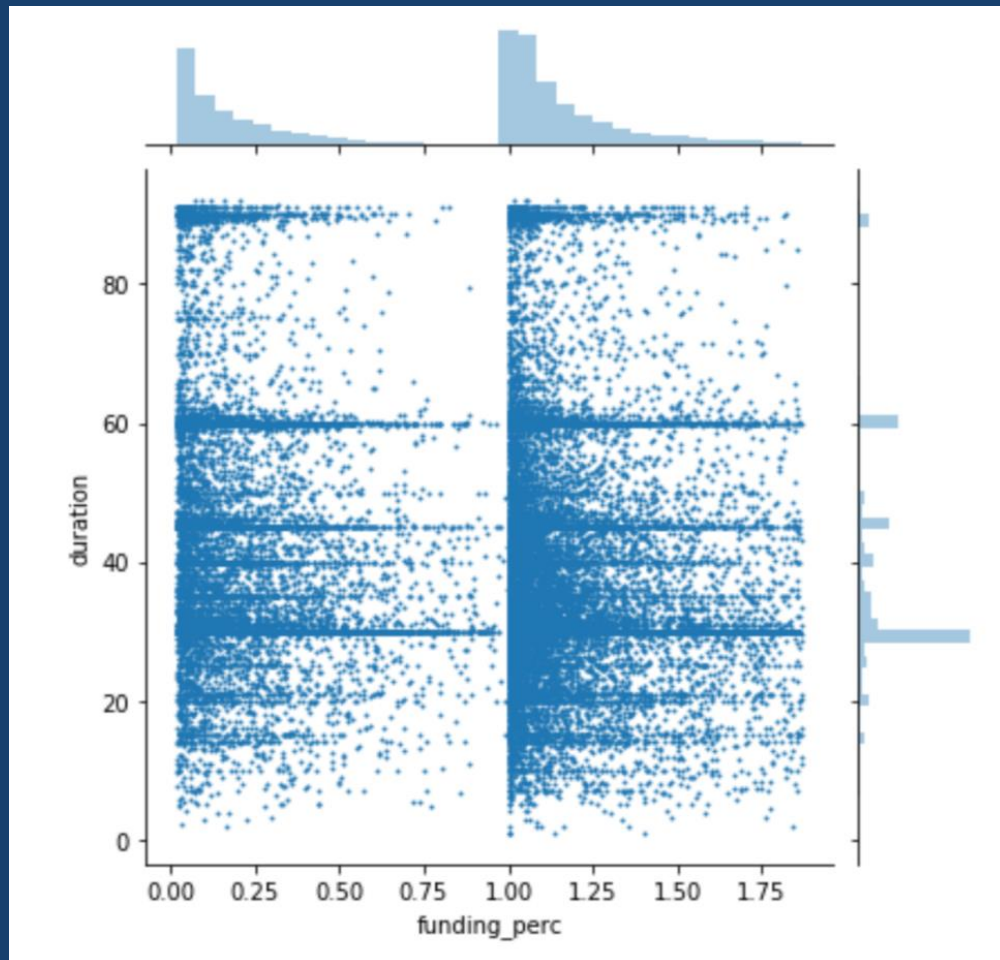
DATA ANALYSIS
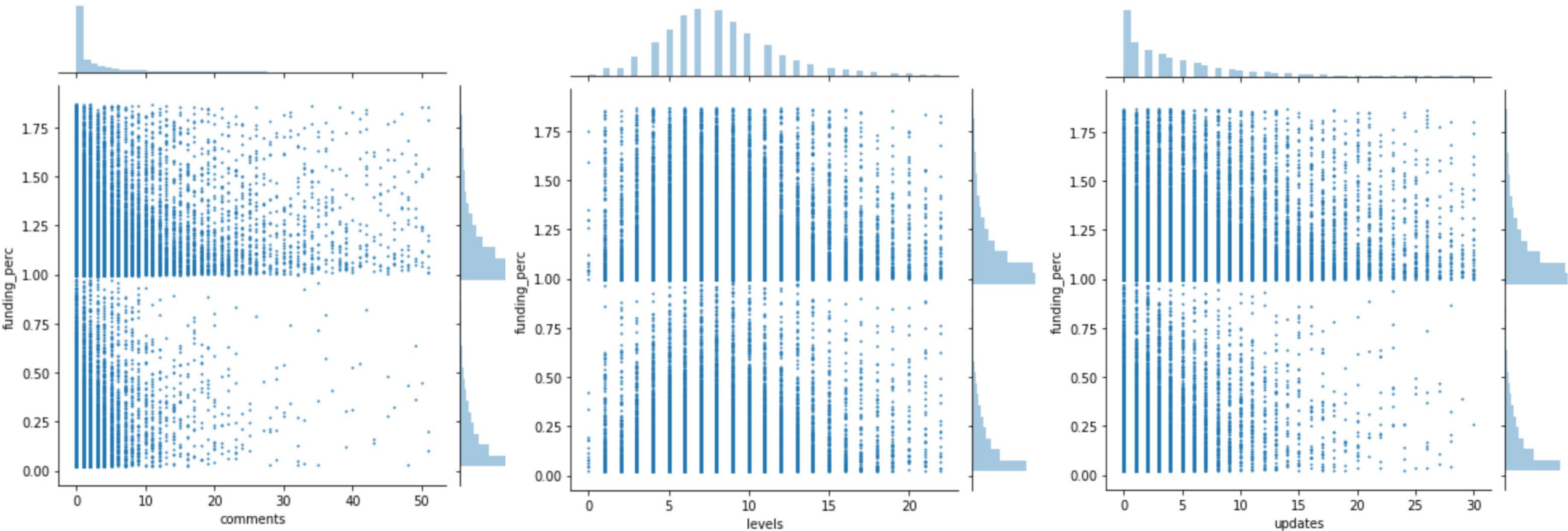## A) MAIN PROJECT CHARACTERISTICS

DATA ANALYSIS
A) MAIN PROJECT CHARACTERISTICS

DATA ANALYSIS
B) ADDITIONAL CHARACTERISTICS

FUTURE DIRECTION

- Obtain more data: at least 200 rows / subcategory

- Make scrapy code more efficient to minimize time taken to scrape

- Build a model to predict the success of a project