

# Python Code for Data Processing

October 27, 2019

```
In [ ]:
```

```
In [ ]: # This python code is to process data for next step analysis
        # we add extra columns such as project duration
```

```
In [ ]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        from scipy import stats
        from sklearn.linear_model import LinearRegression
```

```
In [2]: df = pd.read_csv("Kickstarter_data.csv", header=0, encoding="utf8")
        #open the data set
```

```
In [3]: #check number of records and info
        df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 110815 entries, 0 to 110814
Data columns (total 15 columns):
ID                110815 non-null int64
name              110814 non-null object
category          110815 non-null object
main_category     110815 non-null object
currency          110815 non-null object
deadline          110815 non-null object
goal              110815 non-null int64
launched          110815 non-null object
pledged           110815 non-null float64
state             110815 non-null object
backers           110815 non-null int64
country           110815 non-null object
usd pledged       109999 non-null float64
usd_pledged_real  110815 non-null float64
usd_goal_real     110815 non-null float64
dtypes: float64(4), int64(3), object(8)
memory usage: 12.7+ MB
```

```
In [4]: df.head()
```

```
Out[4]:
```

	ID	name \
0	1015685046	Organic Tattoo remains 10 days (Self use & cus...
1	1019043170	Handcrafted leather wallet   key holder   card...
2	1036288991	The Liaisons: a new twist on the old standards
3	1045749249	The Many Encounters of Bosley Bear
4	1048577059	Swift & Co   Innovative Men's Footwear

  

	category	main_category	currency	deadline	goal	launched \
0	Accessories	Fashion	CAD	1/1/16	6000	12/7/15 18:21
1	Accessories	Fashion	EUR	1/1/16	2000	12/11/15 10:52
2	Jazz	Music	USD	1/1/16	5000	11/9/15 20:12
3	Children's Books	Publishing	USD	1/1/16	20000	11/24/15 2:08
4	Footwear	Fashion	GBP	1/1/16	40000	11/17/15 12:30

  

	pledged	state	backers	country	usd pledged	usd_pledged_real \
0	100.0	failed	2	CA	74.84	70.89
1	2102.0	successful	48	BE	2314.01	2282.97
2	5630.0	successful	116	US	5630.00	5630.00
3	101.0	failed	2	US	101.00	101.00
4	2246.0	failed	19	GB	3416.59	3273.48

  

	usd_goal_real
0	4253.51
1	2172.19
2	5000.00
3	20000.00
4	58298.84

```
In [5]: #number of campaigns under each state
df["state"].value_counts()
```

```
Out[5]: failed      58155
successful    38172
canceled      13178
suspended       707
undefined       596
live              7
Name: state, dtype: int64
```

```
In [6]: df["state"].value_counts(normalize=True)
```

```
Out[6]: failed      0.524794
successful    0.344466
canceled      0.118919
suspended     0.006380
undefined     0.005378
live          0.000063
Name: state, dtype: float64
```

```
In [7]: #check country
df["country"].value_counts()
```

```
Out[7]: US      71905
        GB      11991
        CA      5555
        AU      3181
        DE      2865
        IT      2195
        FR      2084
        ES      1741
        MX      1633
        NL      1124
        SE       982
        N,0"     816
        CH       611
        NZ       607
        DK       597
        HK       577
        SG       523
        IE       466
        BE       465
        AT       443
        NO       378
        LU        47
        JP        29
        Name: country, dtype: int64
```

```
In [8]: #drop unidentified
df = df[df["country"] != 'N,0"]
```

```
In [9]: df["launched"] = pd.to_datetime(df["launched"])    #convert to datetime
```

```
In [10]: df["deadline"] = pd.to_datetime(df["deadline"])  #convert to datetime
```

```
In [11]: df.dtypes
```

```
Out[11]: ID                int64
        name              object
        category          object
        main_category      object
        currency           object
        deadline          datetime64[ns]
        goal              int64
        launched          datetime64[ns]
        pledged           float64
        state             object
        backers           int64
        country           object
```

```

    usd_pledged          float64
    usd_pledged_real     float64
    usd_goal_real        float64
    dtype: object

```

```

In [12]: # calculate the overall fund raising campaign duration
         df["duration"] = (df["deadline"] - df["launched"]).astype('timedelta64[D]') + 1

```

```

In [13]: df.head()

```

```

Out[13]:
      ID                                     name \
0  1015685046  Organic Tattoo remains 10 days (Self use & cus...
1  1019043170  Handcrafted leather wallet | key holder | card...
2  1036288991    The Liaisons: a new twist on the old standards
3  1045749249                The Many Encounters of Bosley Bear
4  1048577059                Swift & Co | Innovative Men's Footwear

      category main_category currency  deadline  goal \
0  Accessories      Fashion    CAD 2016-01-01   6000
1  Accessories      Fashion    EUR 2016-01-01   2000
2      Jazz          Music    USD 2016-01-01   5000
3  Children's Books  Publishing    USD 2016-01-01  20000
4      Footwear      Fashion    GBP 2016-01-01  40000

      launched pledged      state  backers country  usd pledged \
0 2015-12-07 18:21:00   100.0    failed      2      CA      74.84
1 2015-12-11 10:52:00  2102.0  successful     48      BE     2314.01
2 2015-11-09 20:12:00  5630.0  successful    116      US     5630.00
3 2015-11-24 02:08:00   101.0    failed      2      US      101.00
4 2015-11-17 12:30:00  2246.0    failed     19      GB     3416.59

      usd_pledged_real  usd_goal_real  duration
0           70.89      4253.51      25.0
1        2282.97      2172.19      21.0
2        5630.00      5000.00      53.0
3         101.00     20000.00      38.0
4        3273.48     58298.84      45.0

```

```

In [14]: cleaned_df = df.drop(["usd pledged"], axis = 1)

```

```

In [15]: cleaned_df.head()

```

```

Out[15]:
      ID                                     name \
0  1015685046  Organic Tattoo remains 10 days (Self use & cus...
1  1019043170  Handcrafted leather wallet | key holder | card...
2  1036288991    The Liaisons: a new twist on the old standards
3  1045749249                The Many Encounters of Bosley Bear
4  1048577059                Swift & Co | Innovative Men's Footwear

```

	category	main_category	currency	deadline	goal	\
0	Accessories	Fashion	CAD	2016-01-01	6000	
1	Accessories	Fashion	EUR	2016-01-01	2000	
2	Jazz	Music	USD	2016-01-01	5000	
3	Children's Books	Publishing	USD	2016-01-01	20000	
4	Footwear	Fashion	GBP	2016-01-01	40000	

	launched	pledged	state	backers	country	usd_pledged_real	\
0	2015-12-07 18:21:00	100.0	failed	2	CA	70.89	
1	2015-12-11 10:52:00	2102.0	successful	48	BE	2282.97	
2	2015-11-09 20:12:00	5630.0	successful	116	US	5630.00	
3	2015-11-24 02:08:00	101.0	failed	2	US	101.00	
4	2015-11-17 12:30:00	2246.0	failed	19	GB	3273.48	

	usd_goal_real	duration
0	4253.51	25.0
1	2172.19	21.0
2	5000.00	53.0
3	20000.00	38.0
4	58298.84	45.0

In [16]: *#drop the undefined and live campaigns*

```
cleaned_df = cleaned_df[cleaned_df["state"] != 'undefined']
cleaned_df = cleaned_df[cleaned_df["state"] != 'live']
```

In [17]: cleaned\_df["state"].value\_counts()

```
Out[17]: failed      58050
successful    38081
canceled      13157
suspended       705
Name: state, dtype: int64
```

In [18]: cleaned\_df.groupby('main\_category')['category'].value\_counts()

```
Out[18]: main_category  category
Art                  Art      2391
                  Illustration 1387
                  Painting    814
                  Mixed Media  646
                  Public Art   523
                  Digital Art  390
                  Sculpture    370
                  Performance Art 303
                  Conceptual Art 217
                  Installations 179
                  Ceramics     145
                  Textiles     140
                  Video Art     82
```

Comics	Comic Books	1569
	Graphic Novels	1023
	Comics	643
	Webcomics	332
	Anthologies	234
	Events	70
Crafts	Crafts	1620
	Woodworking	521
	DIY	517
	Candles	184
	Stationery	110
	Printing	109
	Crochet	63
	Knitting	55
	Embroidery	52
	Pottery	41
	Glass	40
	...	
Publishing	Calendars	144
	Literary Journals	117
	Comedy	72
	Translations	69
	Literary Spaces	22
	Letterpress	10
Technology	Apps	3067
	Technology	2502
	Gadgets	1516
	Web	1484
	Hardware	1066
	Software	991
	Wearables	655
	DIY Electronics	408
	Sound	348
	3D Printing	254
	Robots	234
	Camera Equipment	214
	Flight	152
	Space Exploration	131
	Fabrication Tools	119
	Makerspaces	103
Theater	Plays	540
	Theater	460
	Musical	363
	Festivals	223
	Experimental	143
	Immersive	129
	Spaces	101
	Comedy	96

Name: category, Length: 170, dtype: int64

```
In [19]: cleaned_df["country"].value_counts(normalize=True)
```

```
Out[19]: US      0.653687
         GB      0.109016
         CA      0.050503
         AU      0.028920
         DE      0.026047
         IT      0.019947
         FR      0.018947
         ES      0.015828
         MX      0.014846
         NL      0.010219
         SE      0.008928
         CH      0.005555
         NZ      0.005519
         DK      0.005428
         HK      0.005246
         SG      0.004755
         IE      0.004237
         BE      0.004218
         AT      0.004028
         NO      0.003437
         LU      0.000427
         JP      0.000264
         Name: country, dtype: float64
```

```
In [20]: # using LabelEncoder we convert the ['state'] column from text to digit
         from sklearn.preprocessing import LabelEncoder
```

```
         le = LabelEncoder()
         le.fit(cleaned_df["state"])
         le_name_mapping = dict(zip(le.classes_, le.transform(le.classes_)))
         print(le_name_mapping)
         cleaned_df["state"] = le.transform(cleaned_df["state"])
```

```
{'canceled': 0, 'failed': 1, 'successful': 2, 'suspended': 3}
```

```
In [21]: # create column 'is_successful' to indicate successful campaign as 1, else 0
         cleaned_df['is_successful'] = cleaned_df['state'].apply(lambda x: 1 if x == 2 else 0)
```

```
In [22]: cleaned_df.head(10)
```

```
Out[22]:
```

	ID	name \
0	1015685046	Organic Tattoo remains 10 days (Self use & cus...
1	1019043170	Handcrafted leather wallet   key holder   card...
2	1036288991	The Liaisons: a new twist on the old standards

```

3 1045749249          The Many Encounters of Bosley Bear
4 1048577059          Swift & Co | Innovative Men's Footwear
5 1049001468          LEMON LIGHTS Debut Album + busking Tour 2016
6 1055979451 Help Will Abele Finish His First Full Length A...
7 1061340742 Dodo Tool. A handy tool that can travel & stay...
8 1067308701 Breathable x Water Proof hoodie w/10+ Practica...
9 1070562691          Jo's Travel Journals

```

```

          category main_category currency  deadline  goal  \
0      Accessories      Fashion      CAD 2016-01-01   6000
1      Accessories      Fashion      EUR 2016-01-01   2000
2           Jazz          Music      USD 2016-01-01   5000
3  Children's Books  Publishing      USD 2016-01-01  20000
4      Footwear          Fashion      GBP 2016-01-01  40000
5      World Music          Music      EUR 2016-01-01   3650
6      Hip-Hop          Music      USD 2016-01-01   2500
7  Product Design      Design      USD 2016-01-01  16000
8      Apparel          Fashion      USD 2016-01-01  25000
9  Literary Journals  Publishing      CAD 2016-01-01   1500

```

```

          launched  pledged  state  backers  country  usd_pledged_real  \
0 2015-12-07 18:21:00   100.00     1         2        CA           70.89
1 2015-12-11 10:52:00  2102.00     2        48        BE          2282.97
2 2015-11-09 20:12:00  5630.00     2       116        US          5630.00
3 2015-11-24 02:08:00   101.00     1         2        US           101.00
4 2015-11-17 12:30:00  2246.00     1        19        GB          3273.48
5 2015-11-03 16:01:00  3750.97     2        24        DE          4073.91
6 2015-12-07 04:34:00   160.00     1         3        US           160.00
7 2015-11-16 12:35:00  2190.00     0        53        US          2190.00
8 2015-11-12 17:16:00 124110.00     2      1093        US         124110.00
9 2015-12-02 22:43:00     0.00     1         0        CA            0.00

```

```

          usd_goal_real  duration  is_successful
0           4253.51         25.0              0
1           2172.19         21.0              1
2           5000.00         53.0              1
3          20000.00         38.0              0
4          58298.84         45.0              0
5           3964.25         59.0              1
6           2500.00         25.0              0
7          16000.00         46.0              0
8          25000.00         50.0              1
9           1063.38         30.0              0

```

```
In [23]: cleaned_df.groupby("usd_goal_real")["goal"].count()
```

```
Out[23]: usd_goal_real
4.900000e-01      1
```



5.500000e-01	1
7.300000e-01	1
7.400000e-01	2
7.500000e-01	3
7.600000e-01	2
7.700000e-01	2
7.800000e-01	1
1.000000e+00	117
1.040000e+00	1
1.060000e+00	2
1.090000e+00	3
1.100000e+00	3
1.110000e+00	1
1.180000e+00	1
1.220000e+00	2
1.230000e+00	1
1.280000e+00	2
1.290000e+00	1
1.300000e+00	1
1.320000e+00	1
1.330000e+00	1
1.410000e+00	1
1.420000e+00	1
1.430000e+00	3
1.450000e+00	1
1.460000e+00	2
2.000000e+00	3
2.240000e+00	1
2.400000e+00	1
...	
1.264302e+07	1
1.405481e+07	1
1.433219e+07	1
1.478743e+07	1
1.520500e+07	1
1.661000e+07	2
1.700000e+07	1
1.710454e+07	1
1.770894e+07	1
1.800000e+07	1
1.913656e+07	1
2.000000e+07	6
2.084109e+07	1
2.200000e+07	1
2.225000e+07	1
2.444817e+07	1
2.500000e+07	3
2.600000e+07	1

2.650000e+07	1
2.853002e+07	1
3.000000e+07	1
3.300000e+07	1
4.000000e+07	3
4.500000e+07	1
5.000000e+07	4
5.164702e+07	1
5.800000e+07	1
7.500000e+07	1
1.000000e+08	12
1.073699e+08	1

Name: goal, Length: 24073, dtype: int64

```
In [24]: # drop the usd_goal_real which is smaller than 1000
cleaned_df[cleaned_df["usd_goal_real"] < 1000].count()
```

```
Out[24]: ID          14868
name          14868
category      14868
main_category  14868
currency      14868
deadline      14868
goal          14868
launched      14868
pledged       14868
state         14868
backers       14868
country       14868
usd_pledged_real 14868
usd_goal_real  14868
duration      14868
is_successful  14868
dtype: int64
```

```
In [25]: cleaned_df.dtypes
```

```
Out[25]: ID          int64
name          object
category      object
main_category  object
currency      object
deadline      datetime64[ns]
goal          int64
launched      datetime64[ns]
pledged       float64
state         int64
backers       int64
country       object
```

```

    usd_pledged_real      float64
    usd_goal_real         float64
    duration              float64
    is_successful         int64
    dtype: object

```

```

In [ ]: # we calculate pledged_ratio to see if the campaign reached the goal
        cleaned_df['pledged_ratio'] = cleaned_df['usd_pledged_real']/cleaned_df['usd_goal_real']
        #cleaned_df['title_len'] = len(cleaned_df['name'])
        cleaned_df['name_length'] = cleaned_df['name'].str.split().str.len()

        #cleaned_df.drop(['title_len'], axis=1)

```

```

In [26]: cleaned_df.sort_values('usd_goal_real', ascending=False)

```

```

Out[26]:
      ID  name \
67491  2132015082  DER NEANDERTALER (Kinofilm mit Starbesetzung i...
38      1279992058  Kybernan Holographic Gaming Network
45651  2137925650  The Time Jumper
14706  1395179504  Bring back Pontiac
99177   941328479  The Throne : Fall of Lucifer
102985  208731425  Animal - The Bear Life Journey
95007   379409641  Peace
2876   1741515953  Let's prove the earth is FLAT! (Suspended) (Su...
30551   813343172  The Castles of the Great Assimil Unearthed (Ca...
26454  1370542565  Art Is Fabulous
37963   843636303  The Exodus, one Ark or many.
59707  1168153740  Sage: Warrior Of The Fallen Gods (Movie)
60470  1119122774  Help me start a fair media company, Trump Media
62910   207256058  Midnight Black
42650  1977067911  Average Joes Journey, To BOLDLY go to the Int...
85457  1227518141  The Life of Alfred Nobel
1858   1180553835  Justyn's
22934   948476327  Batman: Rise of the Cowl
3785   248692941  Atom X Eve - Flawless Entertainment
100801 1857913550  Sorcerer
26907  1834618686  The Cristellianic Castle
14880   51458325  Gouddaa series (Canceled)
8946   477186378  Playboy Mansion Museum! (Canceled)
1814   542785430  Wireless Typewriter (Suspended) (Suspended) (S...
91272  1072735957  Intelligent Design: The Series
36564   647670302  Secure Spent Nuclear Fuel Rods Now
84192   363141153  Palais de la Culture et des Arts
44473   822865000  GamerArchive
25892  2019882363  The Philosophical Degenerate.
28825   440997591  Sparkling Coconut Water Beverage
...      ...      ...
47754  1244531513  Aaron Gibson with Strings Attached

```

97630	2118432531	Angelica Rockne - Queen of San Antonio (Album ...
9182	282731192	The \$1 Poem Project
18484	1317685	Crystal Battery Challenge
105095	465374770	Untitled Album: The Hot Lanes Big Band with Bo...
73885	297840905	Help @Vita become a superstar and comeback his...
32200	636156468	3-D Castle (props for table top RPG and strate...
49013	224199726	Record the new album and shooting a profession...
83606	1767877267	Disidente - The choices we make affect more th...
101158	223030779	Monster Cards! A fun, printable social game fo...
9175	264845956	My Father DIDN'T Kill Himself!
61717	1422679839	The Gateway Persona
96699	104561320	The Other Brothers 'Jones' EP & Merch PREORDER
61876	1694496142	Big Something Tumbleweed Pre-Order
58980	608634146	# KEEP IT CLEAN!
68913	31265352	Oolaf's First Album
67810	1826058000	BACKER FOR LIFE- calendars, cards, posters & m...
107433	641272504	Joe Javascript's Machine Learning Tutorials
48239	1926381309	Smoke Stack Rhino // Love and Other Monsters p...
41114	1202361966	Eyes Without A Face
45845	1739922754	Join the hundreds of backers who are already i...
40432	809291370	Ambient music inspired by Charles Messier
69751	1050732941	LENDEMAIN DE VEILLE
65924	248334240	IRONSAGA
53009	2008375089	Chill-out Digital EP (inspired by Winter) for ...
78038	993961188	SWEET LITTLE BITTER by BAD REED
31208	555587955	Emotional Guide (Canceled)
6037	1793062138	flying cars (Suspended)
83663	69101025	Most percentage funded ever for a Guinness Record
50475	1379346088	Bug's Big Adventure: Mexico Edition

	category	main_category	currency	deadline	goal \
67491	Movie Theaters	Film & Video	EUR	2017-03-14	100000000
38	Video Games	Games	USD	2016-01-01	100000000
45651	Science Fiction	Film & Video	USD	2016-10-10	100000000
14706	Technology	Technology	USD	2016-04-01	100000000
99177	Movie Theaters	Film & Video	USD	2017-10-13	100000000
102985	Painting	Art	USD	2017-11-09	100000000
95007	Film & Video	Film & Video	USD	2017-09-13	100000000
2876	Video	Journalism	USD	2016-01-20	100000000
30551	Fiction	Publishing	USD	2016-06-30	100000000
26454	Art	Art	USD	2016-06-05	100000000
37963	Documentary	Film & Video	USD	2016-08-16	100000000
59707	Fantasy	Film & Video	USD	2017-01-14	100000000
60470	Video	Journalism	USD	2017-01-21	100000000
62910	Art	Art	USD	2017-02-11	75000000
42650	Documentary	Film & Video	USD	2016-09-19	58000000
85457	Television	Film & Video	CHF	2017-07-08	50000000
1858	Music	Music	USD	2016-01-13	50000000

22934	Television	Film & Video	USD	2016-05-16	50000000
3785	Product Design	Design	USD	2016-01-29	50000000
100801	Fantasy	Film & Video	USD	2017-10-26	50000000
26907	Design	Design	USD	2016-06-08	45000000
14880	Television	Film & Video	USD	2016-04-01	40000000
8946	Publishing	Publishing	USD	2016-03-01	40000000
1814	Technology	Technology	USD	2016-01-12	40000000
91272	Science Fiction	Film & Video	USD	2017-08-18	33000000
36564	Television	Film & Video	USD	2016-08-06	30000000
84192	Public Art	Art	EUR	2017-06-30	25000000
44473	Video Games	Games	USD	2016-10-01	26500000
25892	Documentary	Film & Video	USD	2016-06-02	26000000
28825	Drinks	Food	USD	2016-06-19	25000000
...	...	...	...	...	...
47754	Music	Music	USD	2016-10-26	1
97630	Music	Music	USD	2017-10-02	1
9182	Poetry	Publishing	USD	2016-03-02	1
18484	Comics	Comics	USD	2016-04-22	1
105095	Jazz	Music	USD	2017-11-23	1
73885	Pop	Music	USD	2017-04-23	1
32200	Tabletop Games	Games	USD	2016-07-09	1
49013	Pop	Music	USD	2016-11-02	1
83606	Drama	Film & Video	USD	2017-06-26	1
101158	Live Games	Games	USD	2017-10-28	1
9175	Young Adult	Publishing	USD	2016-03-02	1
61717	Rock	Music	USD	2017-02-02	1
96699	Music	Music	USD	2017-09-27	1
61876	Rock	Music	USD	2017-02-03	1
58980	Film & Video	Film & Video	USD	2017-01-07	1
68913	Music	Music	USD	2017-03-23	1
67810	Digital Art	Art	USD	2017-03-16	1
107433	Web	Technology	CAD	2017-12-07	1
48239	Music	Music	AUD	2016-10-29	1
41114	Horror	Film & Video	CAD	2016-09-08	1
45845	Games	Games	CAD	2016-10-12	1
40432	Electronic Music	Music	CAD	2016-09-02	1
69751	Country & Folk	Music	CAD	2017-03-29	1
65924	Video Games	Games	CAD	2017-03-03	1
53009	Electronic Music	Music	CAD	2016-11-29	1
78038	Rock	Music	CAD	2017-05-20	1
31208	Apps	Technology	DKK	2016-07-03	5
6037	Flight	Technology	CAD	2016-02-13	1
83663	Publishing	Publishing	MXN	2017-06-26	10
50475	Art	Art	MXN	2016-11-12	10

		launched	pledged	state	backers	country	\
67491	2017-01-13	20:08:00	36.00	1	2	DE	
38	2015-11-07	00:57:00	13.00	1	4	US	

45651	2016-08-11	12:21:00	0.00	0	0	US
14706	2016-02-01	13:39:00	2.00	1	2	US
99177	2017-08-14	16:45:00	0.00	1	0	US
102985	2017-09-10	17:54:00	59.00	1	2	US
95007	2017-08-14	16:46:00	1.00	1	1	US
2876	2015-12-21	22:29:00	13.00	3	4	US
30551	2016-05-31	21:19:00	0.00	0	0	US
26454	2016-04-06	02:55:00	115.00	1	2	US
37963	2016-06-17	17:20:00	14.00	1	5	US
59707	2016-12-15	02:24:00	0.00	1	0	US
60470	2016-11-22	06:41:00	1.00	1	1	US
62910	2016-12-15	03:08:00	50.00	1	1	US
42650	2016-07-21	18:55:00	21.00	1	5	US
85457	2017-05-09	18:42:00	22.00	1	1	CH
1858	2015-12-14	17:30:00	0.00	1	0	US
22934	2016-03-17	08:17:00	10.00	1	2	US
3785	2015-11-30	21:37:00	10003.00	1	2	US
100801	2017-08-27	22:21:00	271.00	1	4	US
26907	2016-04-09	07:12:00	0.00	1	0	US
14880	2016-02-19	00:52:00	0.00	0	0	US
8946	2016-01-20	08:11:00	0.00	0	0	US
1814	2015-11-13	18:45:00	2.00	3	2	US
91272	2017-07-04	09:00:00	14888.00	1	73	US
36564	2016-07-07	00:08:00	0.00	1	0	US
84192	2017-06-16	09:19:00	130.00	1	2	FR
44473	2016-08-02	04:28:00	6.00	0	6	US
25892	2016-05-03	23:27:00	5.00	1	3	US
28825	2016-05-20	02:01:00	0.00	1	0	US
...	...	...	...	...	...	...
47754	2016-09-26	23:48:00	3285.00	2	39	US
97630	2017-08-04	19:14:00	760.00	2	18	US
9182	2016-02-01	13:13:00	44.00	2	17	US
18484	2016-03-23	19:01:00	75.00	2	12	US
105095	2017-10-25	21:14:00	380.00	2	11	US
73885	2017-03-24	19:18:00	100.00	2	1	US
32200	2016-06-30	22:34:00	150.00	2	4	US
49013	2016-10-18	17:36:00	30.00	2	3	US
83606	2017-05-22	02:45:00	0.00	0	0	US
101158	2017-10-07	20:47:00	153.00	2	96	US
9175	2016-01-20	15:18:00	3431.00	2	155	US
61717	2017-01-03	19:33:00	10.00	2	1	US
96699	2017-08-28	02:10:00	2280.00	2	44	US
61876	2016-12-05	23:34:00	7529.00	2	144	US
58980	2016-12-06	20:58:00	270.00	2	5	US
68913	2017-02-16	16:10:00	687.00	2	16	US
67810	2017-03-09	01:35:00	75.00	2	11	US
107433	2017-11-01	14:41:00	438.14	2	36	CA
48239	2016-10-13	07:35:00	280.00	2	9	AU

41114	2016-07-10 00:38:00	24.00	2	6	CA
45845	2016-09-19 20:16:00	15.00	0	15	CA
40432	2016-08-10 07:34:00	180.00	2	13	CA
69751	2017-01-30 05:12:00	5307.87	2	118	CA
65924	2017-01-31 16:59:00	111.00	2	19	CA
53009	2016-11-05 23:49:00	57.00	2	13	CA
78038	2017-04-20 16:34:00	2000.00	2	42	CA
31208	2016-06-03 14:03:00	0.00	0	0	DK
6037	2016-01-14 17:38:00	0.00	3	0	CA
83663	2017-05-25 19:01:00	9430.00	2	2	MX
50475	2016-11-11 16:30:00	335.00	2	7	MX

	usd_pledged_real	usd_goal_real	duration	is_successful
67491	38.65	1.073699e+08	60.0	0
38	13.00	1.000000e+08	55.0	0
45651	0.00	1.000000e+08	60.0	0
14706	2.00	1.000000e+08	60.0	0
99177	0.00	1.000000e+08	60.0	0
102985	59.00	1.000000e+08	60.0	0
95007	1.00	1.000000e+08	30.0	0
2876	13.00	1.000000e+08	30.0	0
30551	0.00	1.000000e+08	30.0	0
26454	115.00	1.000000e+08	60.0	0
37963	14.00	1.000000e+08	60.0	0
59707	0.00	1.000000e+08	30.0	0
60470	1.00	1.000000e+08	60.0	0
62910	50.00	7.500000e+07	58.0	0
42650	21.00	5.800000e+07	60.0	0
85457	22.72	5.164702e+07	60.0	0
1858	0.00	5.000000e+07	30.0	0
22934	10.00	5.000000e+07	60.0	0
3785	10003.00	5.000000e+07	60.0	0
100801	271.00	5.000000e+07	60.0	0
26907	0.00	4.500000e+07	60.0	0
14880	0.00	4.000000e+07	42.0	0
8946	0.00	4.000000e+07	41.0	0
1814	2.00	4.000000e+07	60.0	0
91272	14888.00	3.300000e+07	45.0	0
36564	0.00	3.000000e+07	30.0	0
84192	148.36	2.853002e+07	14.0	0
44473	6.00	2.650000e+07	60.0	0
25892	5.00	2.600000e+07	30.0	0
28825	0.00	2.500000e+07	30.0	0
...	...	...	...	...
47754	3285.00	1.000000e+00	30.0	1
97630	760.00	1.000000e+00	59.0	1
9182	44.00	1.000000e+00	30.0	1
18484	75.00	1.000000e+00	30.0	1

105095	380.00	1.000000e+00	29.0	1
73885	100.00	1.000000e+00	30.0	1
32200	150.00	1.000000e+00	9.0	1
49013	30.00	1.000000e+00	15.0	1
83606	0.00	1.000000e+00	35.0	0
101158	153.00	1.000000e+00	21.0	1
9175	3431.00	1.000000e+00	42.0	1
61717	10.00	1.000000e+00	30.0	1
96699	2280.00	1.000000e+00	30.0	1
61876	7529.00	1.000000e+00	60.0	1
58980	270.00	1.000000e+00	32.0	1
68913	687.00	1.000000e+00	35.0	1
67810	75.00	1.000000e+00	7.0	1
107433	341.34	7.800000e-01	36.0	1
48239	215.14	7.700000e-01	16.0	1
41114	18.52	7.700000e-01	60.0	1
45845	11.39	7.600000e-01	23.0	0
40432	137.59	7.600000e-01	23.0	1
69751	3978.02	7.500000e-01	58.0	1
65924	82.81	7.500000e-01	31.0	1
53009	42.85	7.500000e-01	24.0	1
78038	1485.22	7.400000e-01	30.0	1
31208	0.00	7.400000e-01	30.0	0
6037	0.00	7.300000e-01	30.0	0
83663	522.81	5.500000e-01	32.0	1
50475	16.41	4.900000e-01	1.0	1

[109993 rows x 16 columns]

```
In [28]: # we calculate pledged_ratio to see if the camapign reached the goal
cleaned_df['pledged_ratio'] = cleaned_df['usd_pledged_real']/cleaned_df['usd_goal_real']
#cleaned_df['title_len'] = len(cleaned_df['name'])
cleaned_df['name_length'] = cleaned_df['name'].str.split().str.len()

#cleaned_df.drop(['title_len'], axis=1)
```

```
In [29]: cleaned_df.head()
```

```
Out[29]:
```

	ID	name \
0	1015685046	Organic Tattoo remains 10 days (Self use & cus...
1	1019043170	Handcrafted leather wallet   key holder   card...
2	1036288991	The Liaisons: a new twist on the old standards
3	1045749249	The Many Encounters of Bosley Bear
4	1048577059	Swift & Co   Innovative Men's Footwear

  

	category	main_category	currency	deadline	goal \
0	Accessories	Fashion	CAD	2016-01-01	6000
1	Accessories	Fashion	EUR	2016-01-01	2000



2	Jazz	Music	USD	2016-01-01	5000
3	Children's Books	Publishing	USD	2016-01-01	20000
4	Footwear	Fashion	GBP	2016-01-01	40000

	launched	pledged	state	backers	country	usd_pledged_real	\
0	2015-12-07 18:21:00	100.0	1	2	CA	70.89	
1	2015-12-11 10:52:00	2102.0	2	48	BE	2282.97	
2	2015-11-09 20:12:00	5630.0	2	116	US	5630.00	
3	2015-11-24 02:08:00	101.0	1	2	US	101.00	
4	2015-11-17 12:30:00	2246.0	1	19	GB	3273.48	

	usd_goal_real	duration	is_successful	pledged_ratio	name_length
0	4253.51	25.0	0	0.016666	10.0
1	2172.19	21.0	1	1.050999	9.0
2	5000.00	53.0	1	1.126000	9.0
3	20000.00	38.0	0	0.005050	6.0
4	58298.84	45.0	0	0.056150	7.0

```
In [30]: cleaned_df.to_csv(r'Kickstarter_data_update2.csv')
```

```
In [ ]:
```