

# Databootcamp Final Project

## Going Beyond the Horn: An Analysis on Unicorn Startups



### Author: Annette Lee

Unicorns are startups valued at over one billion dollars. The startup industry has been a booming one for the last ten years. As the number of startups grow, the valuations have grown to crazy amounts, even if the [product is wildly unsuccessful](#). I thought it would be interesting to analyze the indicators of an unicorn startup and learn more about the characteristics of a billion dollar company. In the end, I hope to answer: What will the next billion-dollar start-up look like?

### Methodology

This analysis is split into two parts. The first is analyzing the data by looking at trends within the majority of startups. The second part is analyzing the

characteristic trends within the most highly valued startups. The three traits I am looking at in both parts are:

### **1. Company Profile**

### **2. Investors**

### **3. Growing Industries**

## **Data Report**

The dataset is about 2016 unicorn startups, high-growth companies valued at over 1 billion dollars. The dataset is from datahub and is a JSON file. It can be accessed here:

[\[https://old.datahub.io/dataset/unicorn-startups-2016\]](https://old.datahub.io/dataset/unicorn-startups-2016)

The dataset gives the following about the 2016 high-valued startups:

*Name* Name of the startup.

*Slogan* Not all startups have one, but it helps to give a mission to the company.

*Description* Startups are infamous for throwing out buzzwords in a pitch as they try to hook investors. It's interesting to see what the buzzwords are for the unicorn startups and if it relates to the overused Silicon Valley jargon that everyone makes fun of.

*Location* Silicon Valley is the most well known startup capital but there are more hotspots emerging as you can see from the other cities on the list.

*Founded* Some startups skyrocket to unicorn status while other take more time.

*Founders* One of the most important factors VCs look at are the founding team because it's believed companies are more successful if there are co-founders.

*Investors* Investors are usually Venture Capital firms, wealthy individuals or public offerings. It's interesting to see who the biggest startup investors are and how much they're investing.

*Employees* The range of number of employees, which indicate firm size.

*Valuation* All above 1 billion dollars, with the highest at \$65 billion.

*Rounds* Rounds are the rounds of financing per year and value amount.

Categories Industries the startup encompasses.

## My Packages

In the analysis below, I will be using the following packages:

- Pandas which will be my core tool to import, manipulate, merge, and analyze the data.
- Numpy which will allow me to perform certain mathematical operations and transformations of the data.
- Matplotlib package which will help me plot my results.
- Urllib which extends my library for opening URLs.
- Collections which allows me to count instances.
- Basemap which will allows me to map certain features of the data.

```
In [2]: import sys                                # system module
import pandas as pd                             # Pandas package
import numpy as np                             # Foundation for
pandas
import matplotlib.pyplot as plt                 # Graphics module
import datetime as dt                          # Date and time m
odule
import urllib.request                           # Library extensi
on for opening URLs
from collections import Counter                 # Specialized Con
tainer module
from mpl_toolkits.basemap import Basemap       # Mapping module
```

## Organizing the Data

Below I walk through the steps to organize the data. I downloaded the data from the source and converted it from a JSON format to Panda using the `pd.read_json` command. The URL was giving me some trouble when I tried accessing it so I used the Urllib package to help me access the site. After the download, I rearranged the columns and cleaned up its contents, in order to make the data easier to manipulate in my analysis.

## Importing the Data

```
In [3]: user_agent = 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_
```

```

1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/63.0.3239.8
4 Safari/537.36'
url= 'https://old.datahub.io/dataset/9bde14f0-d58c-49b8-b758
-dae9edb29dfa/resource/5f6713df-8dad-4202-85ae-8a42eb93c6c6/
download/masterdict.json'
headers={'User-Agent':user_agent,}

request=urllib.request.Request(url,None,headers) #The assemb
led request to open the URL
response = urllib.request.urlopen(request)
data = response.read() # The downloaded data

```

```

In [5]: uni = pd.read_json(data) #Converting from JSON format to Pan
da

```

```

In [6]: uni.head()

```

Out[6]:

	1	2	3	4	
<b>Categories</b>	[Cars, Marketplace, Mobile, Unicorn]	[Hardware, Mobile, Unicorn]	[Travel, Marketplace, Unicorn]	[Big Data, Security, Analytics, Unicorn]	[Me Soc Me Mo Uni
<b>Description</b>	King of ride-hailing apps, Uber now struggles ...	Centering the business around devoted fans, Xi...	Airbnb's online platform lets users rent uniqu...	Palantir sells analytics software for integrat...	Sna pho vide me app
<b>Employees</b>	10k+	1k-5k	1k-5k	1k-5k	500
<b>Founded</b>	2009	2010	2008	2004	201

	1	2	3	4	
<b>Founders</b>	Garrett Camp & Travis Kalanick	Lin Bin, Li Wanqiang, Zhou Guangping, Huang Ji...	Brian Chesky, Joe Gebbia, & Nathan Blecharczyk	Alexander Karp, Stephen Cohen, Peter Thiel, Jo...	Eva Spi Bot Mu

5 rows × 163 columns

After importing, I want to transpose the data so the companies are on the index and the traits on the column heading.

```
In [7]: uni = uni.T
```

```
In [8]: uni.head()
```

```
Out[8]:
```

	Categories	Description	Employees	Founded	Founders	
<b>1</b>	[Cars, Marketplace, Mobile, Unicorn]	King of ride-hailing apps, Uber now struggles ...	10k+	2009	Garrett Camp & Travis Kalanick	KF En As Be G\
<b>2</b>	[Hardware, Mobile, Unicorn]	Centering the business around devoted fans, Xi...	1k-5k	2010	Lin Bin, Li Wanqiang, Zhou Guangping, Huang Ji...	Ac Te Hc Qi Ve
<b>3</b>	[Travel, Marketplace, Unicorn]	Airbnb's online platform lets users rent uniqu...	1k-5k	2008	Brian Chesky, Joe Gebbia, & Nathan Blecharczyk	Se Ca An Hc Ti

	Categories	Description	Employees	Founded	Founders	
4	[Big Data, Security, Analytics, Unicorn]	Palantir sells analytics software for integrat...	1k-5k	2004	Alexander Karp, Stephen Cohen, Peter Thiel, Jo...	Tig Gl Fo Fu AF Ke
5	[Messaging, Social Media, Mobile, Unicorn]	Snapchat's photo and video messaging app with ...	500-1k	2011	Evan Spiegel & Bobby Murphy	KF Lic Be SV Ge

I dropped columns I didn't need and rearranged the columns in a more logical order.

```
In [9]: uni = uni.drop(['Rounds', 'Rank', 'Legal Name'], axis = 1)
# Drop columns
```

```
In [10]: uni.columns
```

```
Out[10]: Index(['Categories', 'Description', 'Employees', 'Founded',
'Founders',
'Investors', 'Location', 'Name', 'Slogan', 'Valuatio
n'],
dtype='object')
```

```
In [11]: column_Titles = ['Name', 'Slogan', 'Description', 'Categories',
'Location', 'Founded', 'Founders', 'Investors', 'Employees',
'Valuation']
uni = uni.reindex(columns=column_Titles)
# Reorder columns
```

```
In [12]: uni.head()
```

```
Out[12]:
```

	Name	Slogan	Description	Categories	Location	Fou
--	------	--------	-------------	------------	----------	-----

	Name	Slogan	Description	Categories	Location	Found
1	Uber	Everyone's Private Driver	King of ride-hailing apps, Uber now struggles ...	[Cars, Marketplace, Mobile, Unicorn]	San Francisco, CA	2009
2	Xiaomi	Only for fans	Centering the business around devoted fans, Xi...	[Hardware, Mobile, Unicorn]	Beijing	2010
3	Airbnb	Belong Anywhere	Airbnb's online platform lets users rent uniqu...	[Travel, Marketplace, Unicorn]	San Francisco, CA	2008
4	Palantir	NaN	Palantir sells analytics software for integrat...	[Big Data, Security, Analytics, Unicorn]	Palo Alto, CA	2004
5	Snapchat	NaN	Snapchat's photo and video messaging app with ...	[Messaging, Social Media, Mobile, Unicorn]	Los Angeles, CA	2011

### Cleaning up data content

There are characters within the content that will make it harder for me to manipulate the data so I cleaned up the data. I focused on the Founders, Investors, Employees and Valuation columns.

### Founders

The founders column has random symbols like & and spaces between the first and last name, which make it hard to convert each name to a string. After the manipulation, each Founder's name is a string.

```
In [13]: uniFOUNDERS = uniFOUNDERS.str.replace(' &', ',').str.replace(',,', ',').str.replace(' ', ',').str.replace(' ', '_')
uniFOUNDERS.str.rsplit(pat=',').head()
```

```
Out[13]: 1      [Garrett_Camp, Travis_Kalanick]
2      [Lin_Bin, Li_Wanqiang, Zhou_Guangping, Huang_J...]
3      [Brian_Chesky, Joe_Gebbia, Nathan_Blecharczyk]
4      [Alexander_Karp, Stephen_Cohen, Peter_Thiel, J...]
5      [Evan_Spiegel, Bobby_Murphy]
Name: Founders, dtype: object
```

### Investors

The Investors column had a similar problem as the Founders, so I cleaned it up the same way I did the Founders column.

```
In [14]: uniINVESTORS = uniINVESTORS.str.replace(' ', ',').str.replace('& ', ',').str.replace(' ', '_')
uniINVESTORS.head()
```

```
Out[14]: 1      KPCB,New_Enterprise_Associates,Benchmark,GV,Fi...
2      Accel,DST,Temasek_Holdings,Qi_Ming_Venture,Mor...
3      Sequoia_Capital,Andreessen_Horowitz,Tiger_Glob...
4      Tiger_Global,Founders_Fund,GSV,ARTIS,Keith_Rab...
5      KPCB,Lightspeed,Benchmark,SV_Angel,General_Cat...
Name: Investors, dtype: object
```

### Valuation

I converted the Valuation column to a float so that the contents became numbers.

```
In [21]: uni.dtypes
```

```
Out[21]: Name          object
Slogan          object
Description      object
Categories       object
```



```

Location      object
Founded       object
Founders      object
Investors     object
Employees     float64
Valuation     object
dtype: object

```

```
In [22]: uni.Valuation = uni.Valuation.astype(float)
```

```
In [23]: uni.head()
```

```
Out[23]:
```

	Name	Slogan	Description	Categories	Location	Found
1	Uber	Everyone's Private Driver	King of ride-hailing apps, Uber now struggles ...	[Cars, Marketplace, Mobile, Unicorn]	San Francisco, CA	2009
2	Xiaomi	Only for fans	Centering the business around devoted fans, Xi...	[Hardware, Mobile, Unicorn]	Beijing	2010
3	Airbnb	Belong Anywhere	Airbnb's online platform lets users rent uniqu...	[Travel, Marketplace, Unicorn]	San Francisco, CA	2008
4	Palantir	NaN	Palantir sells analytics software for integrat...	[Big Data, Security, Analytics, Unicorn]	Palo Alto, CA	2004

	Name	Slogan	Description	Categories	Location	Four
5	Snapchat	NaN	Snapchat's photo and video messaging app with ...	[Messaging, Social Media, Mobile, Unicorn]	Los Angeles, CA	2011

### Categories

It's easier for me to view the categories as separate columns versus as a list. So I created a new column for each word in the Categories column. The new Categories columns can be seen at the far right of the set.

```
In [24]: uni['Categories_001'] = uni['Categories'].str[0]
#Takes the first string in the Categories column and uses it in new column, "Categories_001".
```

```
In [25]: uni['Categories_002'] = uni['Categories'].str[1]
#Repeat for all following strings.
```

```
In [26]: uni['Categories_003'] = uni['Categories'].str[2]
```

```
In [27]: uni['Categories_004'] = uni['Categories'].str[3]
```

```
In [28]: uni['Categories_005'] = uni['Categories'].str[4]
```

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

Out[29]:

	Name	Slogan	Description	Categories	Location	Four
1	Uber	Everyone's Private Driver	King of ride-hailing apps, Uber now struggles ...	[Cars, Marketplace, Mobile, Unicorn]	San Francisco, CA	2009

	Name	Slogan	Description	Categories	Location	Founders
2	Xiaomi	Only for fans	Centering the business around devoted fans, Xi...	[Hardware, Mobile, Unicorn]	Beijing	2010
3	Airbnb	Belong Anywhere	Airbnb's online platform lets users rent uniqu...	[Travel, Marketplace, Unicorn]	San Francisco, CA	2008
4	Palantir	NaN	Palantir sells analytics software for integrat...	[Big Data, Security, Analytics, Unicorn]	Palo Alto, CA	2004
5	Snapchat	NaN	Snapchat's photo and video messaging app with ...	[Messaging, Social Media, Mobile, Unicorn]	Los Angeles, CA	2011

## Analysis

### Part 1

Analyzing the traits a majority of the unicorns possess by looking at:

1. Company Profile
2. Most Active Investors
3. Most Popular Industries

#### 1. Company Profile

The traits I looked at in the company profile are location and the number of founders. The location is interesting to look at because it can indicate where most

of the startups are located. The number of founders is also important because the founding team is critical to a successful startup. One of the most famous incubators, Y-Combinator is known for accepting companies based on the dynamic of the founding team rather than the product itself.

### Location

I counted the number of companies per location using the Counter package I imported at the beginning. I created a separate panda table to look at it more closely.

```
In [30]: uni.Location.head()
```

```
Out[30]: 1    San Francisco, CA
         2         Beijing
         3    San Francisco, CA
         4         Palo Alto, CA
         5         Los Angeles, CA
         Name: Location, dtype: object
```

```
In [31]: loc = Counter(dict(Counter(uni.Location))) #Converted the Location and Counter results to a dicti
```

```
In [32]: locs = pd.DataFrame.from_dict(loc, orient='index').reset_index()
         locs.columns = ['Location', '# of Startups']
         locs = locs.sort_values(by = ('# of Startups'), ascending=False).head(11)
         locs
         #Converted the dictionary to a panda and rearranged the results.
```

```
Out[32]:
```

	Location	# of Startups
0	San Francisco, CA	32
1	Beijing	15
6	New York, NY	15
7	Shanghai	8

	Location	# of Startups
2	Palo Alto, CA	7
17	London	4
13	Berlin	4
23	Hangzhou	3
18	Singapore	3
34	Mountain View, CA	3
36	Redwood City, CA	3

The results are best visualized with a map. I took the top 11 most popular cities and plotted them on a map, with the size of the circle proportioned to the amount of unicorns.

```
In [33]: lon = [-122.4, -73.9, 116.3, 121.5, -122.1, -0.1, 13.4, -122
, 103.8, 120.1, -122.2]
lat = [37.7, 40.7, 39.9, 31.2, 37.5, 51.5, 52.5, 37.3, 1.29,
30.2, 37.4]
crowd = [320, 150, 150, 80, 70, 40, 40, 30, 30, 30, 30]

plt.figure(figsize=(18,8))
m = Basemap(projection='merc',llcrnrlat=-50,urcrnrlat=60,\
            llcrnrlon=-150,urcrnrlon=160,lat_ts=20,resolutio
n='c')
m.drawcoastlines(linewidth=0.25)
m.fillcontinents(color='lightgreen',lake_color='lightskyblue')
# draw parallels and meridians.
m.drawcountries(linewidth=0.25)
m.drawstates(linewidth=0.25)
m.drawmapboundary(fill_color='lightskyblue')
plt.title("Unicorn Locations by Number of Companies", fontsize = 14)
x, y = m(-122.4, 37.7)
plt.plot(x, y, 'ok', markersize=5)
x,y = m(lon, lat)
for x1, y1, c in zip(x, y, crowd):
```

```

# need alpha<1 to get some transparency
# red color is more appropriate
m.plot(x1, y1, 'ro', markersize=c/5, alpha=0.4)

labels = ['San Francisco', 'New York City', 'Beijing', 'Shanghai', 'Palo Alto', 'London', 'Berlin', 'Mountainview', 'Singapore', 'Hangzhou', 'Redwood City']
x_offsets = [10000, -80000, -100000, 1000, 12000, 1000, 14000, -2500000, 1000, 1000, 1000]
y_offsets = [380000, -100000, 15000, -100000, -300000, -450000, -80000, 1000, 1000, -500000, 20000]

for label, xpt, ypt, x_offset, y_offset in zip(labels, x, y, x_offsets, y_offsets):
    plt.text(xpt+x_offset, ypt+y_offset, label)

l1 = plt.scatter([],[], s=320, edgecolors='none')
l2 = plt.scatter([],[], s=150, edgecolors='none')
l3 = plt.scatter([],[], s=80, edgecolors='none')
l4 = plt.scatter([],[], s=70, edgecolors='none')
l5 = plt.scatter([],[], s=40, edgecolors='none')
l6 = plt.scatter([],[], s=30, edgecolors='none')

plt.show()

```



### Location Results

The startup hub locations in the U.S. were unsurprising. I expected to see San Francisco and its greater cities and New York City on the map. But I was surprised with the international results. I didn't realize there was a lot of startup activity in Berlin, which isn't a financial capital like London or Beijing. Also I knew Asia was where a lot of the major valuations were happening, especially with powerhouse companies like Ali Baba and TenCent. However, I was surprised to see Hangzhou because it's not a wellknown city. I was also surprised to see Singapore because it's such a small country and their startup scene isn't as

## Founders

I looked at the number of founders per startup to determine the profile of the founding team. Most founding teams I know are two or more people so I'm curious if the results match my own experience. I used the same commands as before and used Counter then converted to Panda.

```
In [35]: uni.Founders.tail()
```

```
Out[35]: 183      David_Stavens,Mike_Sokolsky,Sebastian_Thrun
        195      Robert_Hohman,Tim_Besse
        204      Vic_Zaud,Samir_Arora,Fernando_Ruarte,Ernie_Cic...
        207      Uber
        214      Alexandra_Wilkis_Wilson,Kevin_Ryan,Alexis_Mayb...
        Name: Founders, dtype: object
```

I made a test dummy to make sure I can count the number of Founders.

```
In [36]: test = uni.Founders.tail(5)
        test
```

```
Out[36]: 183      David_Stavens,Mike_Sokolsky,Sebastian_Thrun
        195      Robert_Hohman,Tim_Besse
        204      Vic_Zaud,Samir_Arora,Fernando_Ruarte,Ernie_Cic...
        207      Uber
        214      Alexandra_Wilkis_Wilson,Kevin_Ryan,Alexis_Mayb...
        Name: Founders, dtype: object
```

```
In [37]: test.str.split(',').str.len()
```

```
Out[37]: 183      3
        195      2
        204      6
```

```

207      1
214      4
Name: Founders, dtype: int64

```

The easiest was creating a new column for the number of founders.

```

In [38]: uni['Number_Founders'] = uni.Founders.str.split(',').str.len
()

```

```

In [39]: uni.head()

```

Out[39]:

	Name	Slogan	Description	Categories	Location	Founders
1	Uber	Everyone's Private Driver	King of ride-hailing apps, Uber now struggles ...	[Cars, Marketplace, Mobile, Unicorn]	San Francisco, CA	2009
2	Xiaomi	Only for fans	Centering the business around devoted fans, Xi...	[Hardware, Mobile, Unicorn]	Beijing	2010
3	Airbnb	Belong Anywhere	Airbnb's online platform lets users rent uniqu...	[Travel, Marketplace, Unicorn]	San Francisco, CA	2008
4	Palantir	NaN	Palantir sells analytics software for integrat...	[Big Data, Security, Analytics, Unicorn]	Palo Alto, CA	2004



	Name	Slogan	Description	Categories	Location	Founders
5	Snapchat	NaN	Snapchat's photo and video messaging app with ...	[Messaging, Social Media, Mobile, Unicorn]	Los Angeles, CA	2011

I created a separate panda table for the number of Founders and the number of companies.

```
In [40]: found = Counter(dict(Counter(uni.Number_Founders)))
```

```
In [41]: found = pd.DataFrame.from_dict(found, orient='index').reset_index()
```

```
In [42]: found.columns= ['#_of_Founders', '#_of_Companies']
```

```
In [43]: found = found.sort_values(by='#_of_Founders', ascending=True)
found
```

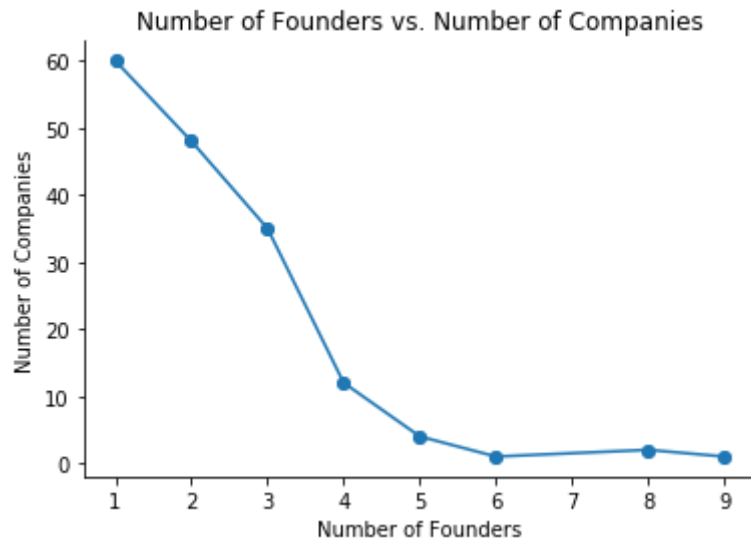
Out[43]:

	#_of_Founders	#_of_Companies
4	1	60
0	2	48
2	3	35
5	4	12
3	5	4
7	6	1
6	8	2
1	9	1

The results are best visualized with a line graph.

```
In [89]: fix, ax = plt.subplots()
ax.scatter(found['#_of_Founders'], found['#_of_Companies'],
           alpha= 0.50)
ax.set_title("Number of Founders vs. Number of Companies")
ax.set_ylabel("Number of Companies")
ax.set_xlabel("Number of Founders")
ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)
plt.plot(found['#_of_Founders'], found['#_of_Companies'], '-o')

plt.show()
```



### ***Number of Founders Results***

Most companies actually have 1-3 founders. Surprisingly, many startups have only 1 founder. There seems to be a rapid decline between 3 and 4 founders.

## **2. Most Active Investors**

I wanted to see who the most active investors were by looking at the most frequent investor. I used the same method as before, Counter then creating a new Panda.

```
In [46]: inv = Counter(dict(Counter(" ".join(uni.Investors).split(',')
)).most_common(22)))
```

```
In [47]: inv = pd.DataFrame.from_dict(inv, orient='index').reset_index()
```

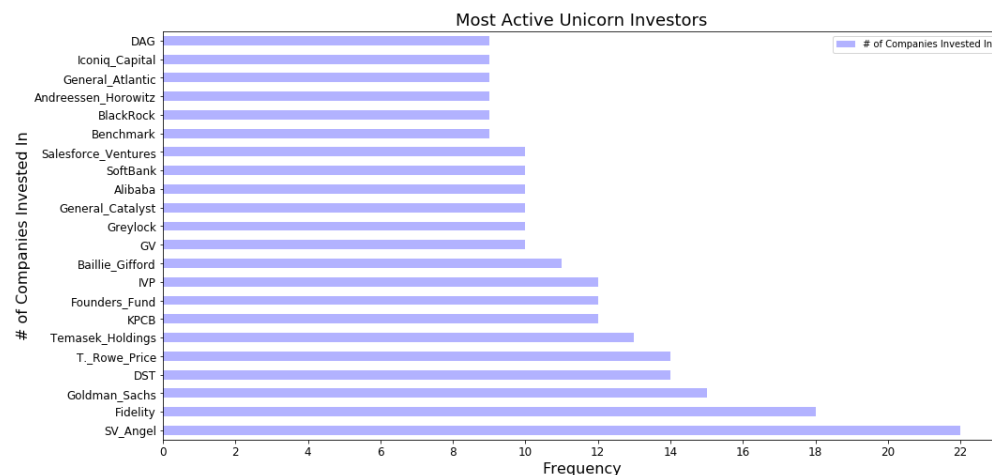
```
In [48]: inv.columns= ['Investor', '# of Companies Invested In']
inv.head()
```

Out[48]:

	Investor	# of Companies Invested In
0	SV_Angel	22
1	Fidelity	18
2	Goldman_Sachs	15
3	DST	14
4	T_Rowe_Price	14

Best visualized with a graph.

```
In [49]: inv.plot(x='Investor', y='# of Companies Invested In', kind=
'barh', xticks = range(0,23,2),color = 'b', stacked = False,
    fontsize = 12, figsize = (16,8), alpha = 0.3)
plt.xlabel('Frequency', fontsize = 16)
plt.ylabel('# of Companies Invested In', fontsize = 16)
plt.title('Most Active Unicorn Investors', fontsize = 18)
plt.show()
```



### Most Active Investors Results

I was familiar with most of the venture capital firms listed above and it was expected they were there because all they do is invest in starts. However, I was surprised to see institutional investors like Fidelity and Temasek Holdings, a Singaporean government investment arm. I was also more surprised to see the banks making as many investments in startups because I thought they focused more on established, less risky companies.

## 3. Growing Industries

The traits I looked at in growing industries are the categories and buzzwords in the description. I wanted to see the industry where most unicorn companies are.

### Popular Industry Categories

I looked at the assigned categories and used Counter and Panda.

```
In [50]: cat1 = Counter(dict(Counter(uni.Categories_001)))
cat2 = Counter(dict(Counter(uni.Categories_002)))
cat3 = Counter(dict(Counter(uni.Categories_003)))
cat4 = Counter(dict(Counter(uni.Categories_004)))
cat5 = Counter(dict(Counter(uni.Categories_005)))
```

I had to sum the counter results from each category.

```
In [51]: cat_sum = cat1 + cat2 + cat3 + cat4 + cat5
```

```
In [52]: cats = pd.DataFrame.from_dict(cat_sum, orient='index').reset_index()
cats.columns = ['Word', 'Frequency']
cats = cats.sort_values(by = ('Frequency'), ascending=False)
cats.head(30)
cats.head()
```

Out[52]:

	Word	Frequency
51	NaN	321
50	Unicorn	163
5	E-Commerce	29
49	Mobile	26
15	Marketplace	19

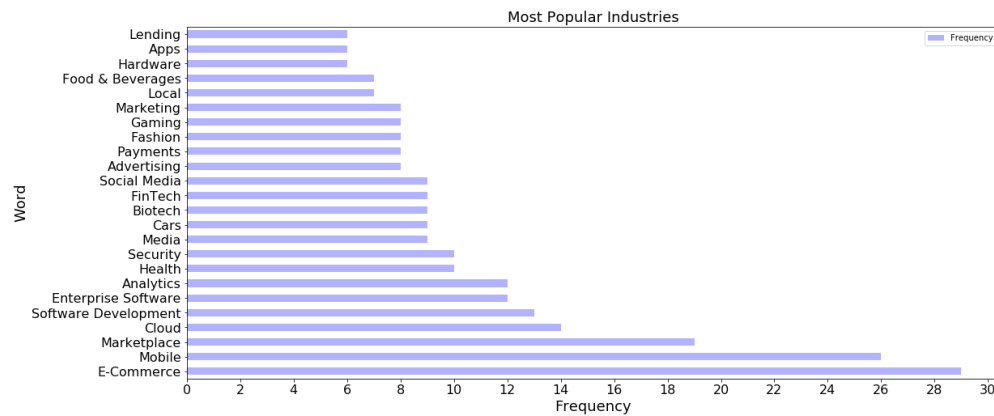
I dropped the NaN and Unicorn because NaN isn't useful and all companies are categorized as unicorns.

```
In [53]: cats = cats.drop([51], axis = 0)
```

```
In [54]: cats = cats.drop([50], axis = 0)
```

```
In [55]: cats = cats.head(24)
```

```
In [56]: cats.plot(x='Word', y='Frequency', kind='barh', xticks = range(0,32,2), color = 'b', stacked = False, fontsize = 16, figsize=(18,8), alpha = 0.3)
plt.xlabel('Frequency', fontsize = 18)
plt.ylabel('Word', fontsize = 18)
plt.title('Most Popular Industries', fontsize = 18)
plt.show()
```



### Categories Results

Unexpectedly, E-Commerce is the most popular industry. It makes sense though, with the growth of Amazon, the decline in brick and mortar retail, and the rise of the direct to consumer model. Also unsurprisingly, Big Data startups that deal with clouds, software development and analytics are also the most frequent.

### Popular Buzzwords

I looked at the buzzwords in the description as another indicator of popular industries. I used the same format as before Count, and Panda.

```
In [57]: uni.Description.head()
```

```
Out[57]: 1    King of ride-hailing apps, Uber now struggles ...
          2    Centering the business around devoted fans, Xi...
          3    Airbnb's online platform lets users rent uniqu...
          4    Palantir sells analytics software for integrat...
          5    Snapchat's photo and video messaging app with ...
          Name: Description, dtype: object
```

Since all the words are in one string, I had to split them by word so I could isolate buzz words.

```
In [58]: uni.Description.str.split(pat=' ').head()
```

```
Out[58]: 1    [King, of, ride-hailing, apps,, Uber, now, str...
          2    [Centering, the, business, around, devoted, fa...
```

```

3   [Airbnb's, online, platform, lets, users, rent...
4   [Palantir, sells, analytics, software, for, in...
5   [Snapchat's, photo, and, video, messaging, app...
Name: Description, dtype: object

```

```

In [59]: results = set()
uni.Description.str.lower().str.split().apply(results.update
)
print (results)
# I split the data into separate words and put them in a tuple.

```

```

{'product', 'shipments', 'asia's', 'curevac', 'revenue', 'brokerage', 'uptake', 'security.', 'gaming.', '15%', 'brand', 'snapdeal', 'than', 'ingestible', 'sparse.', '2,000', 'return', 'request', 'buys', 'banks.', 'presenting', 'boards', 'homes', 'up', 'applied', 'ucar', 'fees.', 'things', 'communication.', 'docker's', 'select', 'chains', 'njoy', 'relatively', 'work', 'listings', 'microsoft', 'save', 'use', 'asia.', 'powa', 'seats', 'defending', 'systems', 'pioneer', 'volume', 'coupa', 'likely', 'wide', 'multiplayer', 'packages.', 'prove', 'lets', 'and', 'flexibility', 'accepts', 'america', 'scenes', 'india's', 'service', 'plans', 'containers', 'check', 'banks', 'maternity', 'airbnb's', '5%', 'appnexus's', 'credit.', 'like.', 'mortgages', 'leading', 'gas', 'market.', 'empty', 'finance', 'quality', 'top', 'client', 'media.', 'coming', 'industry', 'licensing', 'creators', 'rocket', 'draw', 'make', 'music', 'alternative.', 'beibei's', 'red', 'gilt', 'reservations', 'delivery.', 'it', 'plans', 'some', 'expensive', 'launches', 'lenders', 'policy.', 'grabtaxi', 'hotels', 'detection.', 'search', 'utility', 'shazam', 'mid-', 'developer', 'main', 'wifi', 'inspect', 'friendly', 'what', 'zeta', 'hr.', 'recent', 'retail', 'tissues', 'individual', 'performance.', 'comes', 'storing', 'smaller', 'version', 'slims', 'delivering', 'for', 'interoperability', 'commission', 'dinners', 'listings', 'firm', 'pressure', 'readers', 'dominant', 'subscribers.', 'promising', 'help', 'outlet', 'king', 'self-destructing', 'income', 'not', 'hadoop.', 'keeps', 'automating', 'deliver', 'another', 'state', 'hours', 'responsible', 'go', 'territory', 'transaction-ready.', 'fraud', 'originating', 'plus', 'adyen', 'communication', 'commitment', 'rental.', 'health', 'zone', 'merger', 's', 'toward', 'recipe

```

s', 'website,', 'optimize', 'good', 'latin', 'runs', '(mrn  
a)', 'immune', 'offerings', 'airbnb,', 'free', 'square', 'm  
ost', 'mobile-app', 'controlling', 'rates', 'introducing',  
'headsets,', 'supports', 'include', 'holds', 'own', 'geneti  
c', "group's", 'software', 'lending', 'pushing', 'revenu  
e.', 'advantages', 'sharing', 'east.', 'products.', 'transa  
ctions', 'uers', 'systems.', 'headset', 'businesses', 'worl  
d,', 'tips', 'pays', 'tujia', 'beer,', 'personal', 'uses',  
'excellent', 'connect', 'beauty', '500', 'proxies', 'brand  
s', 'pictures', 'subsidiaries', 'insurance,', 'houzz', 'adv  
ertising', 'known', 'human', 'focus', 'existing', 'blood',  
'koudai', 'high-earners', 'week,', 'friendlier', 'audience  
s.', 'profits', 'onto', 'leader', 'generalized', 'includin  
g', 'catalog', 'brands.', 'connected', 'targets', 'loyalt  
y', 'broker,', 'images', 'along', 'sidestepping', 'opportun  
ities.', 'ad', 'restaurants.', 'insurance.', 'peripherals',  
'services,', 'gaming,', 'trackers,', 'store', 'see', 'power  
s', 'does', 'zip', 'zhong', 'security', 'optimizing', 'ente  
rprise', 'corresponding', 'employees,', 'automates', 'desig  
ned,', 'servers,', 'memory', 'speeds.', 'mia.com', 'benefit  
ed', 'rna', 'opportunities', 'medallia', 'analytics.', 'pat  
ients', 'interchange', 'continues', 'implant', 'promises',  
'lakala', 'lifestyle', 'healthy,', 'whitens.', 'child', 're  
ferral', 'wearable', 'didi', 'apps,', 'ventures', 'xiaomi',  
'tickets', 'energy-intensive', 'carpooling,', 'instacart',  
'users.', 'manufactures', 'gambling.', 'established', "snap  
chat's", 'are', 'buy', 'company,', 'self-optimizing', 'casu  
alty', 'except', 'source', 'decision-making', '2006', 'clie  
nts.', 'ones.', 'deem', 'mortar', 'aiming', 'address', 'uni  
que', 'twilio', 'high-performance', 'also', 'jet', 'huge',  
'relying', 'zomato', 'vacation', 'vetted,', 'burning,', 'we  
bsite', 'carrying', 'consumers,', 'lufax', 'china's', 'capt  
ures', 'million', 'subprime', 'this', 'become', 'parentin  
g', 'telemedicine', 'single', 'internationally,', 'celebrit  
y', 'using', 'decolar', 'shoppers', 'features', 'illumio',  
'visual', 'growing', 'bill', 'larger', 'programs', 'autoric  
kshaws', 'things.', 'fanduel', '23andme', 'booking', 'lyf  
t', 'people', 'delivers', 'bots.', 'industries.', 'inmobi',  
'original', 'project', 'e-cigarettes,', 'members,', 'throug  
h', 'group', 'analytic', 'models.', 'struggles', 'spendin  
g,', 'insidesales', 'online', 'controls.', 'them', 'bla',  
'faster', 'amazon', 'behind', 'vine', 'develop', 'autol',  
'variety', 'extensive', 'connections,', 'insurance', 'letti



ng', 'device.', 'mainstay', 'computing', 'shop.', 'relies',  
 'networking', 'bros.', 'designers', '(mrna', 'rapidly', 'mi  
 nus', 'blocked', 'internet', 'which', 'effects', 'trendy',  
 'slower', 'backup.', 'music,', 'southeast', 'restauranteer  
 s.', 'firm,', 'sometimes', 'under', "mongodb's", 'struggl  
 e', 'government.', 'class', 'hero', 'mostly,', 'and,', 'bri  
 ck', 'wepiao', 'adware', 'developers', 'digital', 'companie  
 s.', 'that', 'cloudera', 'solution', 'foodie.', 'prosper',  
 "slack's", 'further', 'b2b', 'therapies', 'goods.', 'shifte  
 d', 'analysis.', 'record', 'channels.', 'molecular', 'gatew  
 ays', 'sprinklr', 'glam', 'anti-virus', 'countries.', 'aid  
 s', 'how', 'event', 'consoles,', "mulesoft's", 'infinidat',  
 "kabam's", 'discussed', 'kind', 'it', 'tech', 'software.',  
 'real-time', "bloom's", 'community', 'service,', 'message  
 s', 'devoted', 'kudu,', 'domo', 'taxis', 'deliveries', 'red  
 undant', 'producers.', 'platform.', 'android', 'app', 'inve  
 ntory', 'themselves.', 'heavyweights,', 'management.', 'fan  
 li', 'young', 'outside', 'difficult-to-destroy', 'be', 'buz  
 zfeed', 'hopes', 'thrives', 'centered', 'one', 'sample', 'c  
 omplaints', 'proteus', 'fans,', 'dji', 'klarna', 'behalf',  
 'major', 'dollars', 'body,', 'home.', 'integrates', 'adtec  
 h', 'east,', 'automatically', 'eyeglasses', 'hosts', 'busin  
 esses,', 'office', 'safety', 'its', 'mouse', 'sites', 'uk-b  
 ased', 'spending', 'code.', 'marketplace.', 'universal', 'y  
 et.', 'ticketing', 'salesforce', 'apis', 'past', 'instituti  
 ons', 'seen', 'different', 'tango', 'asia,', 'primary', 'su  
 ite', 'r&d', 'extremely', 'by', 'questionnaire', 'integrati  
 ng,', 'electricity', 'beer', 'media,', 'without', 'china.',  
 'card', 'boutiques,', 'like', 'mogujie', 'well', 'ship', 's  
 cale', 'wine.', 'china', 'integration.', 'returns.', 'suc  
 h', 'launcher,', 'technical', '20', '100', 'easy-to-use',  
 'would', 'sales', 'buyer,', 'professionals', 'stem', 'lend  
 s', 'relationship', 'many', 'feedback,', 'commission.', 'ac  
 cept', 'difficult', 'technology', 'members-only', 'appdirec  
 t', 'advanced', 'converts', 'set.', 'diabetes', 'centerin  
 g', 'revenues', 'office.', 'autonomous', 'exchange', 'contr  
 ol,', 'marklogic', 'regulation.', 'yelp-like', 'electronic  
 s', 'personalized', 'particularly', 'taobao', 'drawing', 'a  
 ttacks,', 'designs', 'chat.', 'vendor', 'listings.', 'deale  
 rships.', 'scene,', 'demand', 'providing', 'meitu', 'recent  
 ly', 'need', 'messaging,', 'advertising,', 'of', 'draws',  
 'marketing,', 'clients', 'extra', 'jasper', 'efficient,',  
 'data.', 'zscaler', 'blog,', 'discovery,', 'business,', 'sa

vings.', 'treatments', 'wine', 'transactions,', 'they're',  
 'revolutionized', 'platform', 'passes', 'data-driven', 'fea-  
 tures,', 'factors', 'offers', 'note-taking', 'month.', 'des-  
 igner', 'bundled', 'spanning', 'improvement,', 'operates',  
 'accommodations', 'wi-fi', 'efficient', 'lazada', 'dna', 'co-  
 mpeting', 'car.', 'shared', 'passing', 'projects', 'cloudfl-  
 are', 'fees,', 'apptus', 'lookout', 'video-driven', 'youn-  
 g.', 'environments.', 'free-to-play', 'around', 'per', 'ref-  
 erred', 'price', 'chinese', 'uber.', 'well-recognized', 'ty-  
 pically', 'margin,', 'fashion-oriented', 'movie', 'event-  
 s,', 'customizable', 'apron', 'ad-supported', 'benefits',  
 'company', 'hotspots', 'add', 'wireless', 'spacecraft,', 'e-  
 nvironment', 'parts', 'focusing', 'estate', 'moderna', 'i-  
 s', 'cities', 'startups', 'scan', 'high', 'shop', 'number',  
 'offer', 'payments,', 'salaries', 'centers,', 'tencent,',  
 'releases', 'publishers.', 'killing', 'nutritious', 'job',  
 'expanding', 'physical', 'respect.', 'cars', 'designs,', 'c-  
 ompanies,', 'jiuxian', 'mover', 'innovations', 'emerging',  
 "israel's", 'mathematics', 'scanning', 'all', 'point', "ada-  
 ptive's", 'retailer', 'emphasis', 'nutritious,', 'money',  
 'optimization,', 'mobile,', 'security,', 'manages', 'can',  
 'contract', 'trackers.', 'an', 'sports', 'them.', 'gives',  
 'printing', 'avant', 'cheaper.', 'several', 'markups', 'gla-  
 ssdoor', 'cells.', 'quick,', 'industry,', 'filters.', 'crea-  
 tes', 'fees', 'marketing', 'thumbtack', 'in', 'announced',  
 'operating', 'improve', 'generation', 'wework's', 'more',  
 'now', 'palantir', "shazam's", 'format', 'both', 'course',  
 'loans', 'close', 'within', 'experience.', 'there', 'apach-  
 e', 'calls', 'results.', 'video', 'hidden', 'generates', 's-  
 urvey', 'smartphone', 'marketplac', 'business.', 'docusig-  
 n', 'ancestry.', 'unity.', 'access', 'open-source', 'inform-  
 ation.', 'us,', 'fantasy', 'monetizes', 'appdyanmics', 'len-  
 ding,', 'experience', 'rental', 'second', 'diagnostics', 'o-  
 thers', 'classy', 'quikr', 'phones,', 'voice', 'craigslis-  
 t', 'is', 'those', 'deals', 'accessories', 'down.', 'oxfor-  
 d', 'space', 'spacex', 'photo-sharing', 'blue', 'against',  
 'populations', 'apus', 'russia,', 'post', 'items.', 'conten-  
 t', 'backup', '1,000', 'dropbox', 'infant-care', 'lender',  
 'registered', 'series', 'regulatory', 'cruises,', 'dependin-  
 g', 'sites,', 'enterprise,', 'learning', 'machines.', 'sid-  
 e', 'relax', 'shopping', 'rebates.', 'chattiness', 'peak',  
 'neighborhoods.', 'developer', 'chat', 'tool,', 'jawbone',  
 'databases', 'advertising.', 'list', 'legally', 'loans,',

'reducing', 'sale', 'board', 'these', 'file', 'exchange',  
 'clothes', 'unlimited', 'logins', 'crowdsourcing', 'marketp  
 lace', 'systems,', '12', 'airbnb', 'prosper', 'cancer.', 'd  
 istributes', 'companies', 'developer-friendly', 'authenti  
 c', 'purchase', 'mrna', 'rich', 'potentially', 'being', 'im  
 ports,', 'players.', 'your', 'ever,', 'simplivity', 'opport  
 unity', 'yielding', 'only', 'udacity', 'expenses', 'store  
 s', 'low', 'firms,', 'free,', 'throughout', 'low-end', 'app  
 lications,', 'food', 'deal', 'clean', 'combat,', '70+', 'su  
 rveys,', 'result.', 'loans.', 'usable', 'devices.', 'availa  
 ble', 'infrastructure', '2011,', 'blocking', 'devices', 're  
 asonable', 'phones', 'south', 'deals.', 'approach', 'been',  
 'raised', 'hit', 'done', 'drivers', 'benefit', 'bloggers',  
 'targeting', 'allows', 'machine', 'pictures.', "it's", 'ha  
 s', 'regulations,', 'lost', 'appstore', 'ever', 'do', 'orde  
 r', 'happiness.', 'year', 'licensed', 'api', 'data,', 'seri  
 ous', 'entity', 'smooth', 'real', 'alternative', 'swipes,',  
 'android,', 'when', 'rate.', 'shopping', 'immuno-oncology',  
 'mobile', 'ping', 'protein', 'fortune', 'nutanix', 'game  
 s.', 'cost.', 'accuracy', 'malicious', 'records.', 'big',  
 'wearables.', 'markets', 'foreign', 'specializes', 'easie  
 r', 'multiple', 'reuse', 'theaters.', 'much', 'developer  
 s,', 'banking', 'rather', 'reach', 'to', 'send', 'anypoin  
 t', 'attitude', 'options', 'gimbal', 'adds', 'employee', 'c  
 hecks', 'tools', 'concentrated', 'support,', 'documents',  
 'windows,', 'connects', 'otherwise.', 'traffic,', 'secure  
 s', 'eyes', 'e-commerce', 'premium', 'end', 'fun', 'tool  
 s.', 'hardware', 'maker,', 'kabbage', 'adults.', '2008', 'o  
 rders', 'satire', 'surgery', 'provides', 'global', 'borrowe  
 rs.', 'augmented', 'datto', 'fee,', 'publishing,', "spotif  
 y's", 'gaming', 'merchant,', 'putting', 'long', 'warehousin  
 g,', 'come', 'before.', 'selling', 'management', 'taking',  
 'products', 'customers', 'where', 'innovative', 'vintage-in  
 spired', 'try', 'win', 'vehicle', 'signals', 'signing', 'ea  
 rliest', 'gusto', 'service.', 'intelligence', 'cancers.',  
 '70', 'transmission,', 'a', 'sign-in.', 'investors,', 'book  
 marking', 'avast', 'interface', 'few', 'drugs.', 'apps', 'i  
 nformation', 'freemium', 'monitors', 'authenticity', 'aggre  
 ssively', 'reports.', 'drugs,', 'offering', 'email', 'cross  
 -device', 'eight', 'training', 'photography', 'wall', 'aiwu  
 jiwu', 'consumer', 'app,', "company's", 'customer', 'draftk  
 ings', 'cells,', 'form', 'menus', 'surveymonkey', 'user',  
 'mothers', 'from', 'abdominal', 'kinds,', 'enlist', 'cell',

'forms', 'elite', 'sends', 'messaging', 'storage.', 'warne  
r', 'sensors', 'every', 'merchants.', 'china's', 'different  
iate', 'law', 'create', 'alliance', 'video,', 'vice', 'trea  
ts', 'fourth', 'hundreds', 'delivery,', 'uncommon,', 'insti  
tution', 'search,', 'defend', 'increasing', 'cloud-based',  
'across', 'free.', 'monitoring', 'large', 'buying', 'remot  
e', 'reverse', 'developing', 'works.', 'transferwise', 'gre  
ener', 'hollywood,', 'helps', 'checkout', 'hellofresh', 'co  
re', 'fastest', 'massively', 'encrypting', 'alibaba,', 'mes  
senger', 'doctor's', 'opportunity,', 'computers.', 'cance  
r', 'as', 'open', 'retention.', 'explain', 'medicine', 'se  
o,', 'miaopai', 'organized', 'but', 'lamabang', 'created',  
'cut', 'generation,', 'labels', 'cyberattacks.', 'include  
s', 'purchases.', 'ride-hailing', 'password', 'over', 'gith  
ub', 'with', 'credit', 'previously', 'an.', 'tracking', 'de  
stination', 'works', 'in-game', 'intarcia', 'support', 'mas  
sive', 'any', 'qualtrics', 'programs.', 'hadoop', 'early',  
'based', 'content.', 'evernote', 'strengthen', 'riders', 'p  
rimarily', 'communications,', 'risk-management', "organizat  
ion's", 'analyzes', 'microchip', 'moving', 'merchants', 'se  
quencing', 'no', 'vetted', 'media', 'payments', 'medical',  
'chuxing', 'e', 'shops.', 'interactive', 'startup', 'app  
s.', 'request,', 'rate', 'imported', 'capacity', 'ticket',  
'stolen', '30', 'mode', 'high-end', 'year.', '5', 'gourme  
t', 'key', "amazon's", 'employees', 'pinterest', 'improve  
d', 'simplifies', 'allow', 'starting', 'blablabla.', 'logis  
tics', 'lower-mid', 'visualizing,', 'prices.', 'mess', 'cal  
culation.', 'ads', 'avoiding', 'player', 'recovery.', 'smal  
l', 'spotty.', 'vox', 'less', 'about', 'site,', 'applying',  
'rent', 'produces', 'lifestyle,', 'private', 'guaranteein  
g', 'co-produced', 'zenefits', 'high-capacity,', 'other',  
'leap', 'development,', 'server', 'art', 'advice,', 'hailin  
g', 'analyze', 'interpersonal', 'application', 'laws', 'bec  
ause', 'oscar', 'conscious', 'credentials', 'targeted', 'e  
n', 'kik', 'medium', 'let', 'cash-on-delivery', 'dianpin  
g,', 'mongodb.', 'track', 'technologies,', 'alba', 'educati  
on', 'architects', 'optimized', 'sources', 'payments.', 'pr  
omise', 'enterprise.', 'referrals.', 'karma', 'monthly', "m  
ozido's", 'two.', 'still', 'returns', 'screening', 'voic  
e,', 'total', 'daily', 'razer', 'reach.', 'india,', 'son  
g.', 'reports', 'change', 'streaming', 'sellers.', 'copy',  
'wechat,', 'rewards', 'market', 'european', 'bank', 'provid  
er,', 'all-in-one', 'minimalist', 'game', 'take', 'lower',

'market,', 'investors', 'sysadmins', 'carriers,', 'classroom', 'bookings,', 'property', 'nosql', 'furniture.', 'lead s.', 'borrower', 'chauffeured', 'thousands', 'sales,', 'uber', 'sms,', 'flight', 'portable', 'it.', 'cash', 'eventbrite', 'roadtrip', 'services', 'attracting', '3d', 'code,', 'focused', 'student', 'age', 'methods', 'industrial', 'develops', 'prices,', 'ubiquitous', 'discount', 'meituan.com,', 'drives.', 'theranos', 'workers', 'lenders.', 'focuses', 'bio-engineering', 'phone', 'when.', 'lead', 'dining.', 'university,', 'while', 'analyzing', 'takes', 'range', 'everything', 'better', 'volume.', 'retaining', 'photo', "firm's", 'you', 'originally', 'network,', 'websites', 'borrowers', 'itself', 'social', 'trading', 'hyperconverged', 'commerce', 'ingredients', 'expanded', 'then', 'pluralsight', 'share', 'the', 'processes', 'have', 'separate', 'p2p', 'pins', 'publishers,', 'into', 'homemaking,', 'reality', 'database,', 'drone', 'billion', 'storage', 'users', 'provider', 'refinance', 'large-scale', 'wordpress,', 'ground', 'largest', 'grocery', 'beats', 'mostly', 'facebook', 'panshi', 'categories', 'retailer,', 'claims', 'nextdoor', '18-34', 'combines', 'novel', 'energy', 'helping', 'tracks', 'networks.', 'network', 'mall', 'options.', 'welcome', 'courses', 'polymer', 'requiring', 'cloud', 'tests,', 'lines,', 'goes', 'point-of-sale', 'online-only', 'augment', 'retouch', 'control', "customers'", 'reviews', 'recommendations', 'document', 'magic', 'retailers', 'banning', 'point,', 'forum', 'payment', 'each', '10', 'journalism', 'handle', 'institutionalize', 'home,', 'pin', 'films', 'sell', 'hr', "actifio's", 'digitally', 'their', 'international', 'first-mover', 'jessica', 'makes', 'back-end', 'two', 'australia,', 'techniques', 'prices', 'india.', 'brash,', 'despite', 'on', 'charging', 'performance', 'so', 'analytics', 'valuable,', 'receding', 'turnkey', 'viral', 'system,', 'indians', 'or', 'submit', 'local', 'sign', 'videos,', 'easier,', 'smartphones,', 'department', 'street', 'valuable', 'allowing', 'accounts.', 'science', 'farfetch', 'song', 'brand', 'who', 'strength', 'beginning', 'individuals', 'us', 'command', 'mobile.', 'solutions', 'europe,', 'mac.', 'policy,', 'grown', 'employing', 'flipkart', 'meizu', 'reserved', 'anti-terrorism', 'fashion', 'home', 'well-established', 'students', 'making', 'korea', 'popular', 'reviews,', 'system', 'legendary', 'masse,', 'biological', '190+', 'military', 'soon.', 'one97', 'items', 'monetization', 'a,', 'prepare', 'insights', 'satellite

```
s.', 'amasses', 'identifies', 'delivery', 'china,', 'highe
r', 'agency', 'travel', 'ironsource', 'flights,', 'peopl
e,', 'middle', 'immunizations.', 'fitness', 'officially',
'businesses.', 'speakers.', 'buyers', 'tutor', 'first', 'fa
vor', 'platforms,', 'payroll', 'became', "users'", 'distrib
ution', 'maintain', 'productivity', 'new', 'move,', 'fee',
'liquor,', 'installers,', 'billing', 'premier', 'among', 'q
uickly', 'car', 'english', 'pills', 'giving', 'parents,',
'between', 'female', 'wish', 'manufactures,', 'bid', 'shoul
d.', 'seconds,', 'customers,', 'others.', 'restaurant', 'ge
ts', 'handles', 'technology,', 'may', 'at', 'use', 'games',
'consumer,', 'concepts', 'run', 'brazil,', 'skeleton', 'co-
financed', 'qualification', 'cars,', 'natural', 'sized', 'r
ates.', 'keep', 'they', "tanium's", 'records', 'data', 'use
d', 'running', 'staff', 'reporting,', 'successful', 'faste
r,', 'vendors', 'internet,', 'therapy', 'kabam', 'warehouse
s,', 'prescription', 'okta', 'taxi', 'justfab', 'garena',
'foot', 'stories', 'expense', 'makers', 'collaboration.',
'automated', 'business', 'site', 'groceries', 'collects',
'rockets,', 'times', 'bloodwork', 'charges', 'doctors,', 'o
rdering,', 'suppliers', 'compete', 'concessions', 'browse',
'appropriate', 'specialized', 'merchandise.', 'sells', 'off
ered'}
```

```
In [60]: Counter(" ".join(uni.Description).split()).most_common(20)
```

```
Out[60]: [('and', 161),
('a', 89),
('for', 85),
('to', 65),
('of', 61),
('the', 58),
('on', 42),
('in', 39),
('with', 37),
('A', 33),
('from', 31),
('is', 27),
('The', 21),
('offers', 20),
('online', 19),
('software', 17),
('by', 16),
```

```
( 'provides', 16),
( 'company', 15),
( 'its', 15)]
```

I used the Counter but realized that a lot of filler words were the most common. So I wrote a command that ignored these filler words.

```
In [61]: ignore = [ 'at', 'top', 'firm', 'also', 'focuses', 'now', 'lets',
, 'into', 'this', 'the', 'a', 'small', 'local', 'if', 'in', 'it', 'o
f', 'or', 'and', 'for', 'is', 'an', 'to', 'on', 'from', 'wit
h', 'by', 'its', 'offers', 'provides', 'company', 'as', 'ha
s', 'that', 'offering' ]
buzzwords = Counter(" ".join(uni.Description).lower().split
())
for word in list(buzzwords):
    if word in ignore:
        del buzzwords[word]
print (buzzwords)
```

```
Counter({'online': 19, 'software': 17, 'mobile': 15, 'chine
se': 14, 'users': 12, 'data': 11, 'social': 9, 'customer':
9, 'app': 8, 'p2p': 8, 'service': 8, 'payment': 8, 'sells':
7, 'across': 7, 'china's': 7, 'delivery': 7, 'media': 7, 'd
igital': 7, 'platform': 6, 'payments': 6, 'cloud': 6, 'ar
e': 6, 'one': 6, 'originally': 6, 'marketplace': 6, 'sellin
g': 6, 'credit': 6, 'come': 6, 'internet': 6, 'who': 6, 'ra
nge': 6, 'popular': 6, 'business': 5, 'analytics': 5, 'vide
o': 5, 'messaging': 5, 'new': 5, 'services': 5, 'revenues':
5, 'control': 5, 'free': 5, 'up': 5, 'helps': 5, 'focused':
5, 'company,': 5, 'most': 5, 'uses': 5, 'security': 5, 'sal
es': 5, 'many': 5, 'better': 5, 'king': 4, 'other': 4, 'pro
duces': 4, 'million': 4, 'they': 4, 'largest': 4, 'insuranc
e': 4, 'allowing': 4, 'freemium': 4, 'including': 4, 'syste
ms,': 4, 'india,': 4, 'letting': 4, 'providing': 4, 'mediu
m': 4, 'businesses.': 4, 'several': 4, 'middle': 4, 'focusi
ng': 4, 'without': 4, 'then': 4, 'bank': 4, 'network': 4,
'focus': 4, 'marketing': 4, 'connects': 4, 'storage': 4, 'c
an': 4, 'over': 4, 'loans': 4, 'brand': 4, 'charging': 4,
'site': 4, 'apps': 4, 'own': 4, 'outside': 4, 'e-commerce':
4, 'beginning': 4, 'uber': 3, 'around': 3, 'consumer': 3,
'hosts': 3, 'young': 3, 'leading': 3, 'hailing': 3, 'taxi
s': 3, 'than': 3, 'private': 3, 'faster': 3, 'amazon': 3,
'india.': 3, 'shopping': 3, 'per': 3, 'things': 3, 'includ
```

e': 3, 'tools': 3, 'but': 3, 'lenders': 3, 'promises': 3,  
 'smaller': 3, 'making': 3, 'access': 3, 'music': 3, 'high-  
 end': 3, 'sell': 3, 'through': 3, 'market.': 3, 'system': 3,  
 3, 'employees': 3, 'companies': 3, 'data': 3, 'no': 3, 'su  
 ch': 3, 'southeast': 3, 'asia': 3, 'sports': 3, 'hours': 3,  
 3, 'lower': 3, 'buy': 3, 'low': 3, 'human': 3, 'processes':  
 3, 'targeting': 3, 'customers': 3, 'smooth': 3, 'card': 3,  
 'analytics.': 3, 'system': 3, 'advertising': 3, 'mobile': 3,  
 3, 'based': 3, 'car': 3, 'against': 3, 'share': 3, 'their':  
 3, 'fee': 3, 'simplifies': 3, 'which': 3, 'delivers': 3,  
 'reporting': 3, 'presenting': 3, 'profits': 3, 'shop': 3,  
 'what': 3, 'using': 3, 'while': 3, 'prices': 3, 'lifestyl  
 e': 3, 'china': 3, 'early': 3, 'year.': 3, "company's": 3,  
 'them': 3, 'works': 3, 'developers': 3, 'analytic': 3, 'u  
 s': 3, 'health': 3, 'games': 3, 'handles': 3, 'enterprise':  
 3, 'helping': 3, 'borrowers': 3, 'marketing': 3, 'diagnost  
 ics': 3, 'buys': 3, 'all': 3, 'chat': 3, 'businesses': 3,  
 'china': 3, 'food': 3, 'change': 2, 'prepare': 2, 'cars': 2,  
 2, 'systems': 2, 'designs': 2, 'manufactures': 2, 'distri  
 butes': 2, 'countries.': 2, 'analyzing': 2, 'information.': 2,  
 2, 'two': 2, 'product': 2, 'fraud': 2, 'detection.': 2, 'ph  
 oto': 2, 'merger': 2, 'china's': 2, 'taxi': 2, 'rather': 2,  
 'allows': 2, 'tips': 2, 'service.': 2, 'products': 2, 'book  
 marking': 2, 'images': 2, 'pinterest': 2, 'dropbox': 2, 'st  
 orage.': 2, 'competing': 2, 'industry': 2, 'been': 2, 'offi  
 ce': 2, 'businesses': 2, 'valuable': 2, 'targets': 2, 'hug  
 e': 2, 'labels': 2, 'publishers.': 2, 'property': 2, 'provi  
 der': 2, 'big': 2, 'solutions': 2, 'daily': 2, 'expanded':  
 2, 'primary': 2, 'vendors': 2, 'transactions': 2, 'easier':  
 2, 'sign': 2, 'cloudera': 2, 'karma': 2, 'fees.': 2, 'lendi  
 ng': 2, 'internet': 2, 'group': 2, 'successful': 2, 'fashi  
 on': 2, 'markets': 2, 'russia': 2, 'retailer': 2, 'license  
 d': 2, 'headset': 2, 'fitness': 2, 'latin': 2, 'america': 2,  
 2, 'restaurants.': 2, 'makes': 2, 'free-to-play': 2, 'game  
 s': 2, 'player': 2, 'in-game': 2, 'electricity': 2, 'wis  
 h': 2, 'behind': 2, 'items': 2, 'experience.': 2, 'messenger':  
 2, 'rna': 2, 'ubiquitous': 2, 'body': 2, 'carrying':  
 2, 'signals': 2, 'biological': 2, 'tissues': 2, 'cells.': 2,  
 2, 'developing': 2, 'mrna': 2, 'growing': 2, 'communicatio  
 n.': 2, 'developer': 2, 'voice': 2, 'large': 2, 'drivers':  
 2, 'established': 2, 'centered': 2, 'home': 2, 'work': 2,  
 'free': 2, 'ads': 2, 'klarna': 2, 'email': 2, 'pays': 2,  
 'collects': 2, 'recipes': 2, 'ingredients': 2, 'mostly': 2,



```

'github': 2, 'projects': 2, 'open': 2, 'source': 2, 'store
s': 2, 'deliver': 2, 'fees': 2, 'relying': 2, 'referral':
2, 'includes': 2, 'software.': 2, 'holds': 2, 'help': 2, 'l
ending.': 2, 'craigslist': 2, 'cities': 2, 'sellers.': 2,
'automatically': 2, 'monitoring': 2, 'reviews.': 2, 'sharin
g': 2, 'any': 2, 'app.': 2, 'let': 2, 'loyalty': 2, 'progra
ms.': 2, 'specialized': 2, 'nosql': 2, 'wide': 2, 'merchant
s': 2, 'payments.': 2, 'merchants.': 2, 'improve': 2, 'resu
lts.': 2, 'used': 2, 'lead': 2, 'systems.': 2, 'moving': 2,
'insurance.': 2, 'experience': 2, 'giving': 2, 'video.': 2,
'chat.': 2, 'deem': 2, 'operates': 2, 'transactions.': 2,
'travel': 2, 'analyze': 2, 'things.': 2, 'fantasy': 2, 'sit
e.': 2, 'benefits': 2, '2006': 2, 'law': 2, 'banning': 2,
'forms': 2, 'gambling.': 2, 'tools.': 2, 'takes': 2, 'order
s': 2, 'payments.': 2, 'feedback.': 2, 'clients': 2, 'lende
r': 2, 'management': 2, 'reducing': 2, 'need': 2, 'adtech':
2, 'real-time': 2, 'ad': 2, 'premium': 2, 'extremely': 2,
'cloud-based': 2, 'personalized': 2, 'recommendations': 2,
'designs': 2, 'advertising.': 2, 'tech': 2, 'do': 2, 'not':
2, 'handle': 2, 'grown': 2, 'vox': 2, 'runs': 2, 'content':
2, 'genetic': 2, 'china.': 2, 'relax': 2, 'child': 2, 'poli
cy.': 2, 'maternity': 2, 'products.': 2, 'credit.': 2, 'cou
pa': 2, 'integrates': 2, 'search': 2, 'allow': 2, 'program
s': 2, 'hardware': 2, 'users.': 2, 'mall': 2, 'internationa
l': 2, 'good': 2, 'parts': 2, 'gaming': 2, 'movie': 2, 'ano
ther': 2, 'enterprise.': 2, 'enterprise.': 2, 'promising':
2, 'sized': 2, 'technical': 2, 'quality': 2, 'does': 2, 'in
teractive': 2, 'information': 2, 'restaurant': 2, 'deals':
2, 'cancer': 2, 'companies.': 2, 'smartphone': 2, '10': 2,
'about': 2, 'avant': 2, 'seats': 2, 'dna': 2, 'treatments':
2, 'distribution': 2, 'uptake': 2, 'business.': 2, 'printin
g': 2, 'retailer.': 2, 'red': 2, 'learning': 2, 'educatio
n': 2, 'job': 2, 'ride-hailing': 1, 'apps.': 1, 'struggle
s': 1, 'regulations.': 1, 'autonomous': 1, 'develop': 1, 'l
ogistics': 1, 'industries.': 1, 'centering': 1, 'devoted':
1, 'fans.': 1, 'xiaomi': 1, 'phones': 1, 'electronics': 1,
'lower-mid': 1, 'market': 1, 'asia.': 1, 'airbnb's': 1, 're
nt': 1, 'unique': 1, 'accomodations': 1, '190+': 1, 'palant
ir': 1, 'integrating.': 1, 'visualizing.': 1, 'lines.': 1,
'anti-terrorism': 1, "snapchat's": 1, 'self-destructing':
1, 'messages': 1, 'became': 1, 'hit': 1, 'people.': 1, 'med
ia.': 1, 'companies.': 1, 'didi': 1, 'chuxing': 1, 'cars':
1, 'extra': 1, 'india's': 1, 'dominant': 1, 'destination':

```

```

1, '20': 1, '70+': 1, 'categories': 1, '5': 1, 'shipments':
1, 'month.': 1, 'spacex': 1, 'launches': 1, 'advanced': 1,
'rockets.': 1, 'spacecraft.': 1, 'satellites.': 1, 'visua
l': 1, 'tool.': 1, 'people': 1, 'pin': 1, 'boards': 1, 'sav
e': 1, 'pictures': 1, 'like.': 1, 'some': 1, 'pins': 1, 'bu
ying': 1, 'opportunity.': 1, 'yielding': 1, 'revenue.': 1,
'well-recognized': 1, 'first-mover': 1, 'heavyweights.': 1,
'introducing': 1, 'photo-sharing': 1, 'security.': 1, 'wewo
rk's': 1, 'shared': 1, 'space': 1, 'startups': 1, 'expensiv
e': 1, 'square': 1, 'foot': 1, 'networking': 1, 'opportunit
ies.': 1, 'lufax': 1, 'support': 1, 'major': 1, 'borrower
s.': 1, 'theranos': 1, 'technology': 1, 'blood': 1, 'draw
s': 1, 'medical': 1, 'tests.': 1, 'bloodwork': 1, 'easie
r.': 1, 'faster.': 1, 'cheaper.': 1, 'music': 1, 'streamin
g': 1, 'internationally.': 1, "spotify's": 1, 'gives': 1,
'catalog': 1, 'passing': 1, 'cut': 1, 'dji': 1, 'drone': 1,
'technology.': 1, 'flight': 1, 'photography': 1, 'platform
s.': 1, 'gimbal': 1, 'transmission.': 1, 'remote': 1, 'cont
rols.': 1, 'first': 1, 'online-only': 1, 'casualty': 1, 'zh
ong': 1, 'automated': 1, 'insurance.': 1, 'created': 1, 'al
ibaba.': 1, 'tencent.': 1, 'ping': 1, 'an.': 1, 'deal': 1,
'website.': 1, 'snapdeal': 1, 'compete': 1, 'flipkart': 1,
'developer-friendly': 1, 'options.': 1, 'zenefits': 1, 'dif
ferent': 1, 'kind': 1, 'broker.': 1, 'hr': 1, 'department':
1, 'revenue': 1, 'comes': 1, 'when': 1, 'policy.': 1, 'runn
ing': 1, 'open-source': 1, 'apache': 1, 'hadoop.': 1, 'rece
ntly': 1, 'announced': 1, 'kudu.': 1, 'hadoop': 1, 'alterna
tive.': 1, 'consumer.': 1, 'checks': 1, 'hidden': 1, 'firm
s.': 1, 'generates': 1, 'borrower': 1, 'leads.': 1, 'comin
g': 1, 'rocket': 1, 'cmbines': 1, 'retailers': 1, 'emergin
g': 1, 'brazil.': 1, 'east.': 1, 'officially': 1, 'merchand
ise.': 1, 'starting': 1, 'military': 1, 'combat.': 1, 'jawb
one': 1, 'shifted': 1, 'wearable': 1, 'technologies.': 1,
'minimalist': 1, 'trackers.': 1, 'headsets.': 1, 'speaker
s.': 1, 'europe.': 1, 'australia.': 1, 'where': 1, 'chain
s': 1, 'uncommon.': 1, 'hero': 1, 'machine': 1, 'zone': 1,
'massively': 1, 'multiplayer': 1, 'items.': 1, "bloom's":
1, 'energy': 1, 'server': 1, 'converts': 1, 'natural': 1,
'gas': 1, 'burning.': 1, 'rates': 1, 'greener': 1, 'energy-
intensive': 1, 'shoppers': 1, 'fun': 1, 'prices.': 1, 'scen
es.': 1, 'these': 1, 'ship': 1, 'only': 1, 'vehicle': 1, 'w
all': 1, 'street': 1, 'investors': 1, 'hollywood.': 1, 'leg
endary': 1, 'co-produced': 1, 'co-financed': 1, 'series':

```

1, 'films': 1, 'warner': 1, 'bros.': 1, 'universal': 1, 'pictures.': 1, '(mrna)': 1, 'moderna': 1, 'drugs.': 1, 'peak': 1, '2011.': 1, 'core': 1, 'brand.': 1, 'income': 1, 'adults.': 1, 'try': 1, 'clothes': 1, 'delivery.': 1, 'slacks': 1, 'suite': 1, 'fastest': 1, 'b2b': 1, 'application': 1, 'ever.': 1, 'friendly': 1, 'alternative': 1, 'mess': 1, 'uk-based': 1, 'powa': 1, 'point': 1, 'sale': 1, 'peripherals': 1, 'swipes.': 1, 'back-end': 1, 'tracking': 1, 'game': 1, 'garena': 1, 'communications.': 1, 'network.': 1, 'accounts.': 1, 'inmobi': 1, 'optimizing': 1, 'us.': 1, 'lyft': 1, 'become': 1, 'difficult': 1, 'differentiate': 1, 'much': 1, 'larger': 1, 'except': 1, 'receding': 1, 'commitment': 1, 'friendlier': 1, 'riders': 1, 'autorickshaws': 1, 'throughout': 1, 'defending': 1, 'uber.': 1, 'vice': 1, 'itself': 1, 'premier': 1, 'outlet': 1, 'everything': 1, 'satire': 1, 'serious': 1, 'journalism': 1, 'attitude': 1, 'command': 1, 'respect.': 1, 'improvement.': 1, 'houzz': 1, 'architects': 1, 'designers': 1, 'browse': 1, 'concepts': 1, 'charges': 1, 'referrals.': 1, 'checkout': 1, 'address': 1, 'zip': 1, 'code.': 1, 'merchant.': 1, 'minus': 1, 'buyer.': 1, 'aids': 1, 'blue': 1, 'apron': 1, 'nutritious.': 1, 'gourmet': 1, 'dinners': 1, 'two.': 1, 'south': 1, 'korea': 1, 'deliveries': 1, 'customers.': 1, 'legally': 1, 'blocked': 1, 'storing': 1, 'domo': 1, 'automates': 1, 'intelligence': 1, 'drawing': 1, 'hundreds': 1, 'sources': 1, 'customizable': 1, 'reports.': 1, 'facebook': 1, 'developers.': 1, 'collaboration.': 1, 'platform.': 1, 'delivering': 1, 'groceries': 1, 'instacart': 1, 'grocery': 1, 'car.': 1, 'markups': 1, 'shop.': 1, 'magic': 1, 'leap': 1, 'plans': 1, 'augmented': 1, 'reality': 1, 'onto': 1, 'users': 1, 'eyes': 1, 'would': 1, 'see': 1, 'otherwise.': 1, 'yet.': 1, 'female': 1, 'fashion-oriented': 1, 'fees.': 1, 'mogujie': 1, 'extensive': 1, 'marketplace.': 1, 'nutanix': 1, 'innovative': 1, 'centers': 1, 'techniques': 1, 'previously': 1, 'reserved': 1, 'players.': 1, 'performance': 1, 'goes': 1, 'go': 1, 'down.': 1, 'surveys.': 1, 'surveymonkey': 1, 'questionnaire': 1, 'survey': 1, 'trendy': 1, '2,000': 1, 'eight': 1, 'brands.': 1, 'plans.': 1, 'one97': 1, 'mobile.': 1, 'struggle': 1, 'indians': 1, 'use': 1, 'cash': 1, 'be': 1, 'spotty.': 1, 'mover': 1, 'prosper': 1, 'lost': 1, 'ground': 1, '2008': 1, 'regulatory': 1, 'expanding': 1, 'novel': 1, 'territory': 1, 'surgery': 1, 'community': 1, 'lenders.': 1, 'format': 1, 'prosper': 1, 'there': 1, 'because': 1, 'populations':

```

1, 'concentrated': 1, 'retail': 1, 'offerings': 1, 'spars
e.': 1, 'registered': 1, 'intarcia': 1, 'bio-engineering':
1, 'abdominal': 1, 'implant': 1, 'releases': 1, 'diabetes':
1, 'medicine': 1, "tanium's": 1, 'approach': 1, 'quick,':
1, 'easy-to-use': 1, "organization's": 1, 'computers.': 1,
'doctors,': 1, 'bookings,': 1, 'filters.': 1, 'conscious':
1, 'consumers,': 1, 'particularly': 1, 'parents,': 1, 'reli
es': 1, 'jessica': 1, 'alba': 1, 'promise': 1, 'healthy,':
1, 'responsible': 1, 'happiness.': 1, 'docusign': 1, 'docum
ent': 1, 'contract': 1, 'signing': 1, 'those': 1, 'send':
1, 'documents': 1, 'others': 1, 'digitally': 1, 'device.':
1, "asia's": 1, 'booking': 1, 'grabtaxi': 1, 'laws': 1, 're
quest': 1, 'lakala': 1, 'point-of-sale': 1, 'readers,': 1,
'utility': 1, 'bill': 1, 'services,': 1, 'database,': 1, "m
ongodb's": 1, 'security,': 1, 'support,': 1, 'training': 1,
'mongodb.': 1, 'global': 1, 'adyen': 1, 'accept': 1, 'risk-
management': 1, 'viral': 1, 'stories': 1, 'original': 1, 'b
uzzfeed': 1, 'science': 1, 'draw': 1, 'audiences.': 1, "fir
m's": 1, 'productivity': 1, 'sales,': 1, 'insidesales': 1,
"israel's": 1, 'adware': 1, 'scene,': 1, 'ironsource': 1,
'state': 1, 'art': 1, 'installers,': 1, 'typically': 1, 'bu
ndled': 1, 'shopping,': 1, 'koudai': 1, 'apps.': 1, 'genera
tion': 1, 'commission.': 1, "mulesoft's": 1, 'anypoint': 1,
'apis': 1, 'interoperability,': 1, 'efficient': 1, 'reuse':
1, 'interchange': 1, 'between': 1, 'applied': 1, 'mathemati
cs': 1, 'institutionalize': 1, 'data-driven': 1, 'decision-
making': 1, 'fortune': 1, '500': 1, 'clients.': 1, 'opportu
nities': 1, 'oscar': 1, 'unlimited': 1, 'telemedicine': 1,
'trackers.': 1, 'calls': 1, 'smartphones,': 1, 'tango': 1,
'rich': 1, 'environment': 1, 'available': 1, 'you': 1, 'ind
ividual': 1, 'dining.': 1, 'track': 1, 'spending,': 1, 'adv
antages': 1, 'jasper': 1, 'wireless': 1, 'connections,': 1,
'cell': 1, 'carriers,': 1, 'billing': 1, 'connected': 1, 'd
evices.': 1, 'draftkings': 1, 'version': 1, 'note-taking':
1, 'evernote': 1, 'augment': 1, 'memory': 1, 'earliest': 1,
'sites,': 1, 'fanduel': 1, 'benefited': 1, 'lazada': 1, 'ac
cepts': 1, 'cash-on-delivery': 1, 'returns.': 1, 'servic
e,': 1, 'medallia': 1, 'complaints': 1, 'retention.': 1, 'h
igh-earners': 1, 'refinance': 1, 'student': 1, 'loans,': 1,
'finance': 1, 'treats': 1, 'members,': 1, 'applying': 1, 'i
nterpersonal': 1, 'pressure': 1, 'keep': 1, 'mortgages': 1,
'personal': 1, 'loans.': 1, "actifio's": 1, 'copy': 1, 'red
undant': 1, 'file': 1, 'welcome': 1, 'innovations': 1, "cus

```

tomers': 1, 'them.': 1, 'appnexus's': 1, 'main': 1, 'exhan  
 ge,': 1, 'buyers': 1, 'bid': 1, 'wordpress,': 1, 'powers':  
 1, 'fourth': 1, 'scale': 1, 'infinidat': 1, 'efficient,':  
 1, 'high-capacity,': 1, 'high-performance': 1, 'reasonabl  
 e': 1, 'cost.': 1, 'sprinklr': 1, 'brands': 1, 'how': 1, "t  
 hey're": 1, 'being': 1, 'discussed': 1, 'networks.': 1, 'cl  
 assy': 1, 'vintage-inspired': 1, 'prescription': 1, 'eyegla  
 sses': 1, 'price': 1, 'avoiding': 1, 'store': 1, 'expense  
 s': 1, 'licensing': 1, 'appdyanmics': 1, 'monitors': 1, 'ma  
 nages': 1, 'large-scale': 1, 'government.': 1, 'ticketing':  
 1, 'startup': 1, 'events,': 1, 'eventbrite': 1, 'tickets':  
 1, 'insights': 1, 'event': 1, 'producers.': 1, 'justfab':  
 1, 'sends': 1, 'subscribers.': 1, 'manufactures': 1, 'know  
 n': 1, 'celebrity': 1, 'models.': 1, 'demand': 1, 'side':  
 1, 'trading': 1, 'cross-device': 1, 'nextdoor': 1, 'organiz  
 ed': 1, 'neighborhoods.': 1, 'may': 1, 'prove': 1, 'valuabl  
 e,': 1, 'monetization': 1, 'still': 1, 'works.': 1, 'despit  
 e': 1, "doctor's": 1, 'advice,': 1, 'patients': 1, 'take':  
 1, 'pills': 1, 'should.': 1, 'putting': 1, 'ingestible': 1,  
 'microchip': 1, 'drugs,': 1, 'proteus': 1, 'record': 1, 'wh  
 en.': 1, 'twilio': 1, 'voice,': 1, 'sms,': 1, 'api': 1, 'co  
 mmunication,': 1, 'depending': 1, 'volume.': 1, 'single':  
 1, 'blog,': 1, 'sites': 1, '18-34': 1, 'age': 1, 'set.': 1,  
 'submit': 1, 'sample': 1, '23andme': 1, 'reports': 1, 'you  
 r': 1, 'likely': 1, 'ancestry.': 1, 'offered': 1, 'past':  
 1, 'hopes': 1, 'return': 1, 'screening': 1, 'therapy': 1,  
 'soon.': 1, 'android': 1, 'launcher,': 1, 'apus': 1, 'targe  
 ted': 1, 'marketplac': 1, 'e': 1, 'autol': 1, 'individual  
 s': 1, 'dealerships.': 1, 'mainstay': 1, 'industry,': 1, 'a  
 vast': 1, 'anti-virus': 1, 'windows,': 1, 'android,': 1, 'm  
 ac.': 1, "beibei's": 1, 'thrives': 1, 'guaranteeing': 1, 's  
 afety': 1, 'authenticity': 1, 'sidestepping': 1, 'lends':  
 1, 'class': 1, 'workers': 1, 'have': 1, 'cloudflare': 1, 'r  
 everse': 1, 'proxies': 1, 'websites': 1, 'defend': 1, 'cybe  
 rattacks.': 1, "it's": 1, 'servers,': 1, '5%': 1, 'traffi  
 c,': 1, 'so': 1, 'higher': 1, 'speeds.': 1, 'tracks': 1, 'e  
 mployee': 1, 'spending': 1, 'records.': 1, 'purchase': 1,  
 'request,': 1, 'expense': 1, 'vendor': 1, 'relationship':  
 1, 'agency': 1, 'decolar': 1, 'options': 1, 'flights,': 1,  
 'hotels,': 1, 'cruises,': 1, 'vacation': 1, 'packages.': 1,  
 "docker's": 1, 'containers': 1, 'run': 1, 'environments.':  
 1, 'sysadmins': 1, 'fanli': 1, 'rebates.': 1, 'farfetch':  
 1, 'elite': 1, 'boutiques,': 1, 'brick': 1, 'mortar': 1, 's

hops.': 1, 'relatively': 1, 'investors.': 1, 'returns': 1,  
 'rates.': 1, 'european': 1, 'home.': 1, 'lifestyle.': 1, 'f  
 urniture.': 1, 'bloggers': 1, 'others.': 1, 'illumio': 1,  
 'secures': 1, 'computing': 1, 'encrypting': 1, 'controllin  
 g': 1, 'every': 1, 'point.': 1, 'physical': 1, 'machines.':  
 1, "kabam's": 1, 'make': 1, 'purchases.': 1, 'kabam': 1, 'k  
 ik': 1, 'requiring': 1, 'phone': 1, 'number': 1, 'monetize  
 s': 1, 'bots.': 1, 'primarily': 1, 'a.': 1, 'ad-supported':  
 1, 'forum': 1, 'mothers': 1, 'parenting': 1, 'homemaking.':  
 1, 'lamabang': 1, 'imported': 1, 'lookout': 1, 'stolen': 1,  
 'phones.': 1, 'blocking': 1, 'attacks.': 1, 'scanning': 1,  
 'malicious': 1, 'code.': 1, 'more': 1, 'features': 1, 'mark  
 logic': 1, 'databases': 1, 'flexibility.': 1, 'search.': 1,  
 'development.': 1, 'backup.': 1, 'provider': 1, 'panshi':  
 1, 'optimization.': 1, 'pluralsight': 1, 'video-driven': 1,  
 'courses': 1, 'it.': 1, 'students': 1, 'monthly': 1, 'fee':  
 1, 'content.': 1, 'qualtrics': 1, 'captures': 1, 'amasses':  
 1, 'usable': 1, 'institutions': 1, 'quikr': 1, 'claims': 1,  
 '12': 1, 'listings': 1, '30': 1, '1,000': 1, 'listings.':  
 1, 'generation.': 1, 'mouse': 1, 'optimized': 1, 'gaming.':  
 1, 'razer': 1, 'accessories': 1, 'consoles.': 1, 'messagin  
 g.': 1, 'wearables.': 1, 'hyperconverged': 1, 'all-in-one':  
 1, 'simplivity': 1, 'improved': 1, 'capacity': 1, 'performa  
 nce.': 1, "shazam's": 1, 'identifies': 1, 'song': 1, 'few':  
 1, 'seconds.': 1, 'user': 1, 'opportunity': 1, 'song.': 1,  
 'taking': 1, 'shazam': 1, 'keeps': 1, 'free.': 1, 'transfer  
 wise': 1, 'money': 1, 'commission': 1, 'exchange': 1, 'rat  
 e.': 1, 'airbnb.': 1, 'tujia': 1, 'control.': 1, 'employin  
 g': 1, 'staff': 1, 'clean': 1, 'inspect': 1, 'homes': 1, 'w  
 ell': 1, 'maintain': 1, 'accuracy': 1, 'listings.': 1, '7  
 0': 1, 'subsidiaries': 1, 'kinds.': 1, 'aiming': 1, 'streng  
 th': 1, 'unity.': 1, 'gaming.': 1, 'zeta': 1, 'analyzes':  
 1, 'attracting': 1, 'retaining': 1, 'existing': 1, 'ones.':  
 1, 'menus': 1, 'zomato': 1, 'discovery.': 1, 'ordering': 1,  
 1, 'reservations.': 1, 'restauranteers.': 1, 'firm.': 1, 'z  
 scaler': 1, 'recent': 1, 'meituan.com.': 1, 'dianping.': 1,  
 'yelp-like': 1, 'business.': 1, 'plus': 1, 'alliance': 1,  
 'strengthen': 1, 'both': 1, 'therapies': 1, 'killing': 1,  
 'stem': 1, 'cells.': 1, 'difficult-to-destroy': 1, 'creator  
 s': 1, 'cancers.': 1, 'ucar': 1, 'mobile-app': 1, 'chauffeu  
 red': 1, 'originating': 1, 'rental': 1, 'maker.': 1, 'meiz  
 u': 1, 'mid-': 1, 'low-end': 1, 'market.': 1, 'young.': 1,  
 'makers': 1, 'order': 1, 'total': 1, 'volume': 1, 'increasi

ng': 1, '15%': 1, 'year': 1, 'hellofresh': 1, 'correspondin  
 g': 1, 'each': 1, 'week,': 1, 'nutritious': 1, 'excellent':  
 1, 'home.': 1, 'developer': 1, 'meitu': 1, 'creates': 1, 'f  
 eatures,': 1, 'beauty': 1, 'retouch': 1, 'slims': 1, 'white  
 ns.': 1, "mozido's": 1, 'commerce': 1, 'banks.': 1, 'suppor  
 ts': 1, 'rewards': 1, 'deals.': 1, 'like': 1, 'banking': 1,  
 'automating': 1, 'qualification': 1, 'calculation.': 1, 'su  
 bprime': 1, 'leader': 1, 'roadtrip': 1, 'carpooling,': 1,  
 'empty': 1, 'long': 1, 'drives.': 1, 'rate': 1, 'chattines  
 s': 1, 'bla': 1, 'blablabla.': 1, 'ticket': 1, 'wechat,':  
 1, 'wepiao': 1, 'concessions': 1, 'select': 1, 'theaters.':  
 1, 'salesforce': 1, 'apptus': 1, 'close': 1, 'interface':  
 1, 'microsoft': 1, 'office.': 1, 'oxford': 1, 'universit  
 y,': 1, 'revolutionized': 1, 'sequencing': 1, 'molecular':  
 1, 'analysis.': 1, 'portable': 1, 'devices': 1, 'rapidly':  
 1, 'scan': 1, 'protein': 1, 'within': 1, 'reach.': 1, 'je  
 t': 1, 'beats': 1, "amazon's": 1, 'less': 1, 'variety': 1,  
 'slower': 1, 'delivery,': 1, 'suppliers': 1, 'warehousin  
 g,': 1, 'aggressively': 1, 'potentially': 1, 'explain': 1,  
 'project': 1, 'done': 1, 'thumbtack': 1, 'list': 1, 'profes  
 sionals': 1, 'it.': 1, '(mrna': 1, ')is': 1, 'curevac': 1,  
 'immunizations.': 1, 'pioneer': 1, 'e-cigarettes,': 1, 'njo  
 y': 1, 'form': 1, 'factors': 1, 'channels.': 1, 'continue  
 s': 1, 'r&d': 1, 'appropriate': 1, 'regulation.': 1, 'okt  
 a': 1, 'logins': 1, 'gateways': 1, 'applications,': 1, 'pay  
 roll': 1, 'gusto': 1, 'brokerage,': 1, 'hr.': 1, 'develop  
 s': 1, 'infrastructure': 1, 'industrial': 1, 'sensors': 1,  
 'offer': 1, 'optimize': 1, "adaptive's": 1, 'check': 1, 'be  
 nefit': 1, 'immuno-oncology': 1, 'enlist': 1, 'immune': 1,  
 'cancer.': 1, 'sometimes': 1, 'referred': 1, 'airbnb': 1,  
 'aiwujiwu': 1, 'real': 1, 'estate': 1, 'rental.': 1, 'appdi  
 rect': 1, 'appstore': 1, '3d': 1, 'polymer': 1, 'high': 1,  
 'end': 1, 'use,': 1, '100': 1, 'times': 1, 'ever': 1, 'see  
 n': 1, 'before.': 1, 'datto': 1, "'s": 1, 'backup': 1, 'rec  
 overy.': 1, 'turnkey': 1, 'solution': 1, 'data.': 1, 'win  
 e': 1, 'beer': 1, 'jiuxian': 1, 'imports,': 1, 'warehouse  
 s,': 1, 'beer,': 1, 'liquor,': 1, 'and,': 1, 'mostly,': 1,  
 'authentic': 1, 'wine.': 1, 'operating': 1, 'banks,': 1, 'k  
 abbage': 1, 'quickly': 1, 'records': 1, 'mia.com': 1, 'taob  
 ao': 1, 'infant-care': 1, 'foreign': 1, 'goods.': 1, 'vin  
 e': 1, 'miaopai': 1, 'create': 1, 'second': 1, 'videos,':  
 1, 'add': 1, 'effects': 1, 'result.': 1, 'well-establishe  
 d': 1, 'emphasis': 1, 'seo,': 1, 'ventures': 1, 'gets': 1,

```
'behalf': 1, 'client': 1, 'transaction-ready.': 1, 'wifi':
1, 'skeleton': 1, 'key': 1, 'uers': 1, 'connect': 1, 'wi-f
i': 1, 'hotspots': 1, 'password': 1, 'sign-in.': 1, 'crowds
ourcing': 1, 'credentials': 1, 'among': 1, 'east,': 1, 'spa
nning': 1, 'multiple': 1, 'methods': 1, 'english': 1, 'inst
itution': 1, 'world,': 1, 'tutor': 1, "group's": 1, 'self-o
ptimizing': 1, 'classroom': 1, 'management.': 1, 'generaliz
ed': 1, 'massive': 1, 'course': 1, 'udacity': 1, 'specializ
es': 1, 'vetted,': 1, 'designed,': 1, 'glassdoor': 1, 'vett
ed': 1, 'reviews': 1, 'salaries': 1, 'employees,': 1, 'pos
t': 1, 'themselves.': 1, 'board': 1, 'pushing': 1, 'towar
d': 1, 'further': 1, 'integration.': 1, 'mode': 1, 'advertisi
ng,': 1, 'publishing,': 1, 'thousands': 1, 'publishers,':
1, 'glam': 1, 'media,': 1, 'brash,': 1, 'foodie.': 1, 'sepa
rate': 1, 'entity': 1, 'win': 1, 'favor': 1, 'reach': 1, 'm
ove,': 1, 'raised': 1, 'billion': 1, 'dollars': 1, 'under':
1, 'members-only': 1, 'website': 1, 'designer': 1, 'discoun
t': 1, 'prices,': 1, 'gilt': 1, 'inventory': 1, 'en': 1, 'm
asse,': 1, 'adds': 1, 'margin,': 1, 'passes': 1, 'along':
1, 'savings.': 1})
```

I converted the results to a panda data frame.

```
In [62]: buzz = pd.DataFrame.from_dict(buzzwords, orient='index').res
et_index()
buzz.columns = ['Word', 'Frequency']
buzz.head()
```

Out[62]:

	Word	Frequency
0	king	4
1	ride-hailing	1
2	apps,	1
3	uber	3
4	struggles	1

I'm going to take the top 16 ranked words since there are a lot of ties.

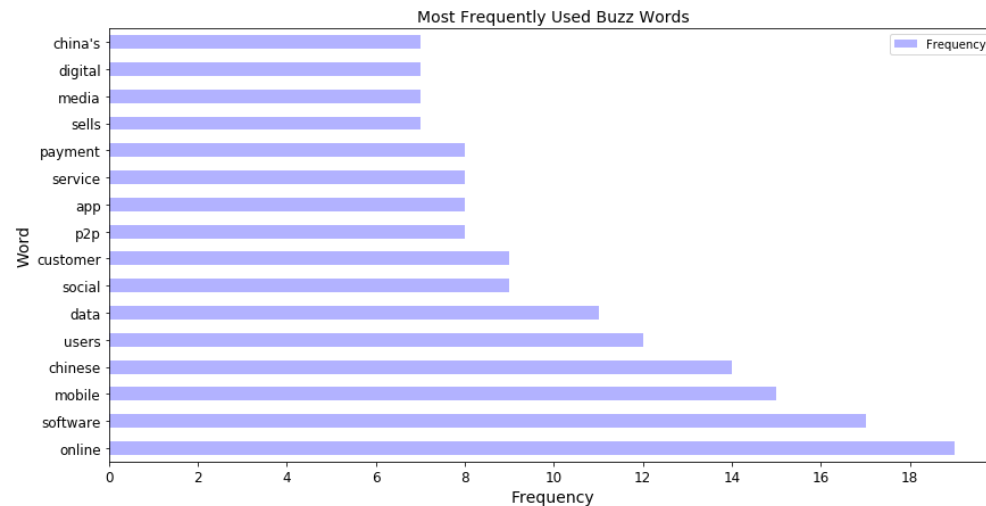


```
In [63]: buzz = buzz.sort_values(by = ('Frequency'), ascending=False)
        .head(16)
        buzz = buzz.reset_index()
        buzz.head()
```

Out[63]:

	index	Word	Frequency
0	31	online	19
1	43	software	17
2	318	mobile	15
3	378	chinese	14
4	33	users	12

```
In [64]: buzz.plot(x='Word', y='Frequency', kind='barh', xticks = range(0,20,2),color = 'b', stacked = False, fontsize = 12, figsize = (14,7), alpha = 0.3)
        plt.xlabel('Frequency', fontsize = 14)
        plt.ylabel('Word', fontsize = 14)
        plt.title('Most Frequently Used Buzz Words', fontsize = 14)
        plt.show()
```



## Part II

The second part of the analysis is comparing the same traits as before but with the valuation of the startup. I want to see if there is a difference between the trait results of the majority of unicorns versus the most highly valued startups.

## 1. Company Profile

Again, I will look at location and number of founders but instead by the most valued startups, not frequency.

### Location

I used the groupby command to group the companies by location. The average valuation was calculated for each location. I then took the top 10 locations with the highest average valuations.

```
In [65]: newloc = (uni.groupby("Location").mean().sort_values(by = 'valuation', ascending=False))
newloc = newloc.head(10)
```

```
In [66]: newloc = newloc.reset_index()
newloc
```

Out[66]:

	Location	Employees	Valuation	Number_Founders
0	Los Angeles, CA	750.0	16.000000	2.000000
1	Bangalore	NaN	15.000000	2.000000
2	Hawthorne, CA	NaN	12.000000	1.000000
3	Shenzhen	NaN	8.000000	1.000000
4	Beijing	3000.0	6.666667	1.600000
5	Palo Alto, CA	3000.0	5.942857	2.428571
6	Stockholm	NaN	5.350000	2.500000
7	San Francisco, CA	6500.0	5.234375	2.468750
8	South San Francisco, CA	NaN	5.000000	2.000000

	Location	Employees	Valuation	Number_Founders
9	Dania, FL	NaN	4.500000	1.000000

```
In [80]: fig = plt.figure()
lon = [-118.2, 77, -118.4, 114, 116.3, -122.1, 18, -122.4, -122.4, -80]
lat = [34, 13, 34, 22.5, 40, 37.5, 59, 37.7, 37.7, 26]
crowd = [160, 150, 120, 80, 66.7, 59.4, 53.5, 52.3, 50, 45]

plt.figure(figsize=(18,8))
m = Basemap(projection='merc',llcrnrlat=-10,urcrnrlat=65,
            llcrnrlon=-150,urcrnrlon=160,lat_ts=20,resolutio
n='c')
m.drawcoastlines(linewidth=0.25)
m.fillcontinents(color='lightgreen',lake_color='lightskyblue')
# draw parallels and meridians.
m.drawcountries(linewidth=0.25)
m.drawstates(linewidth=0.25)
m.drawmapboundary(fill_color='lightskyblue')
plt.title("Unicorn Locations by Valuation", fontsize = 14)
x, y = m(-122.4, 37.7)
plt.plot(x, y, 'ok', markersize=5)
x,y = m(lon, lat)
for x1, y1, c in zip(x, y, crowd):
    # need alpha<1 to get some transparency
    # red color is more appropriate
    m.plot(x1, y1, 'ro', markersize=c/5, alpha=0.4)

labels = ['Greater LA', 'Bangalore', '', 'Shenzhen','Beijing', 'Palo Alto', 'Stockholm', 'Greater SF', '', 'Dania, FL']
x_offsets = [-10000, -80000, -100000, 1000, 12000, 1000, 14000, -250000,1000, 1000, 1000]
y_offsets = [-300000, -100000, 15000, -100000, -300000, -450000, -80000, 1000, 1000, -500000, 20000]

for label, xpt, ypt, x_offset, y_offset in zip(labels, x, y, x_offsets, y_offsets):
    plt.text(xpt+x_offset, ypt+y_offset, label)
```

```

lon = [-122.4, -73.9, 116.3, 121.5, -122.1, -0.1, 13.4, -122
, 103.8, 120.1, -122.2]
lat = [37.7, 40.7, 39.9, 31.2, 37.5, 51.5, 52.5, 37.3, 1.29,
30.2, 37.4]
crowd = [320, 150, 150, 80, 70, 40, 40, 30, 30, 30, 30]

plt.figure(figsize=(18,8))
m = Basemap(projection='merc',llcrnrlat=-50,urcnrlat=60,\
            llcrnrlon=-150,urcnrlon=160,lat_ts=20,resolutio
n='c')
m.drawcoastlines(linewidth=0.25)
m.fillcontinents(color='lightgreen',lake_color='lightskyblu
e')
# draw parallels and meridians.
m.drawcountries(linewidth=0.25)
m.drawstates(linewidth=0.25)
m.drawmapboundary(fill_color='lightskyblue')
plt.title("Unicorn Locations by Number of Companies", fontsi
ze = 14)
x, y = m(-122.4, 37.7)
plt.plot(x, y, 'ok', markersize=5)
x,y = m(lon, lat)
for x1, y1, c in zip(x, y, crowd):
    # need alpha<1 to get some transparency
    # red color is more appropriate
    m.plot(x1, y1, 'ro', markersize=c/5, alpha=0.4)

labels = ['San Francisco', 'New York City', 'Beijing', 'Shan
ghai','Palo Alto', 'London', 'Berlin', 'Mountainview', 'Sing
apore', 'Hangzhou', 'Redwood City']
x_offsets = [10000, -80000, -100000, 1000, 12000, 1000, 1400
00, -2500000,1000, 1000, 1000]
y_offsets = [380000, -100000, 15000, -100000, -300000, -4500
00, -80000, 1000, 1000, -500000, 20000]

for label, xpt, ypt, x_offset, y_offset in zip(labels, x, y,
x_offsets, y_offsets):
    plt.text(xpt+x_offset, ypt+y_offset, label)

l1 = plt.scatter([],[], s=320, edgecolors='none')

```

```

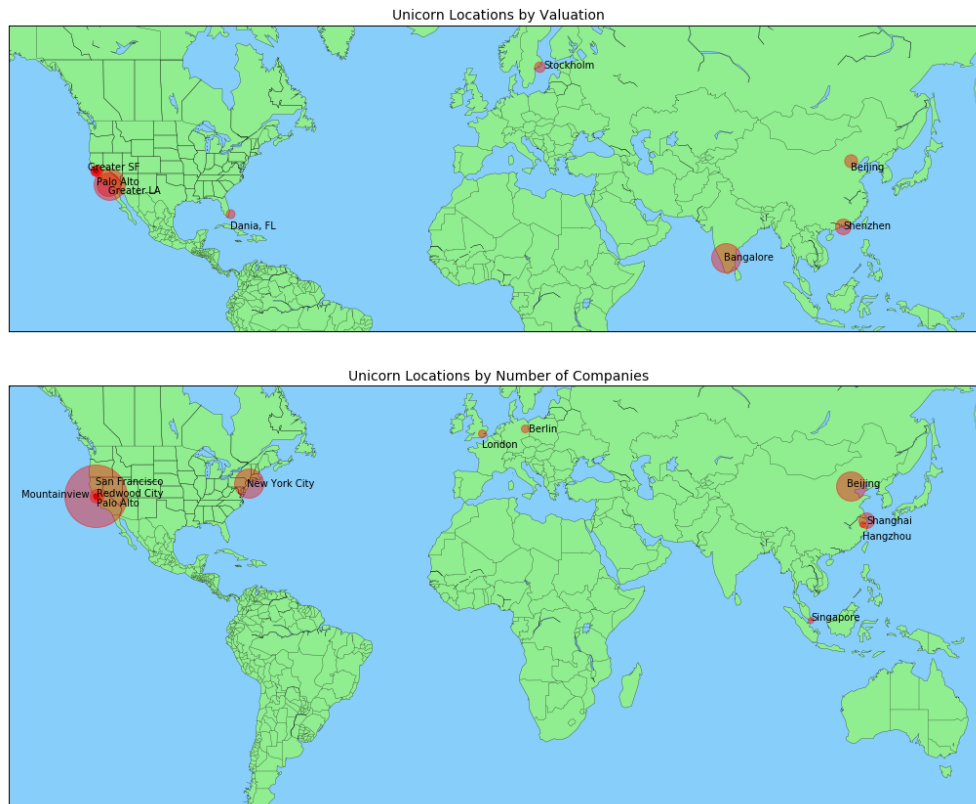
12 = plt.scatter([],[], s=150, edgecolors='none')
13 = plt.scatter([],[], s=80, edgecolors='none')
14 = plt.scatter([],[], s=70, edgecolors='none')
15 = plt.scatter([],[], s=40, edgecolors='none')
16 = plt.scatter([],[], s=30, edgecolors='none')

```

```

plt.show()
<matplotlib.figure.Figure at 0x111472240>

```



### Location Results

The results are interesting because they vary greatly from the map of where most of the startups are located. The new additions to the map are Los Angeles, Bangalore, Dania, FL, Shenzhen, and Sweden. This is very surprising because I thought there would be more overlap between the locations of most valued and most number of startups. I also don't traditionally see Los Angeles, Florida, or Stockholm as major hubs of startup activity. The one thing that did stay constant was the high valuations in San Francisco/ Greater San Francisco area. Although

the startup landscape is changing, the scene is still the biggest at its birthplace of

## Founders

I used groupby to arrange the Number of Founders by their average valuation.

```
In [70]: f = (uni.groupby("Number_Founders").mean())
f = f.reset_index ()
```

```
In [71]: f
```

```
Out[71]:
```

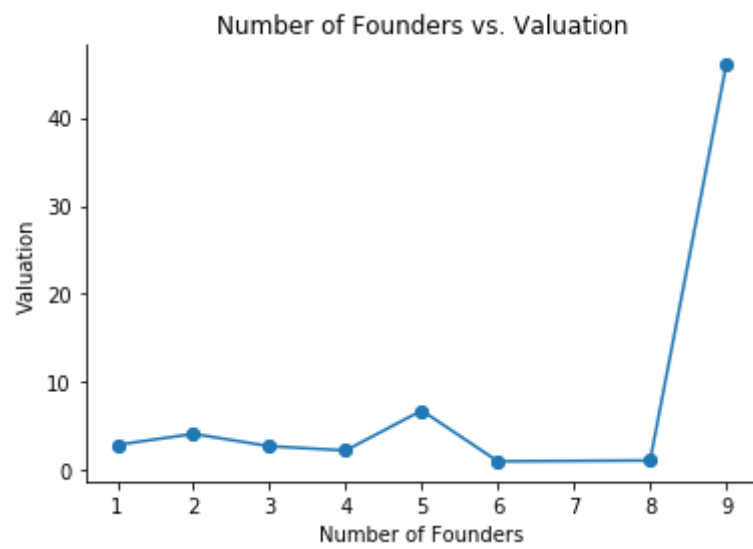
	Number_Founders	Employees	Valuation
0	1	NaN	2.873333
1	2	5375.0	4.116667
2	3	3000.0	2.722857
3	4	NaN	2.233333
4	5	3000.0	6.775000
5	6	NaN	1.000000
6	8	NaN	1.100000
7	9	3000.0	46.000000

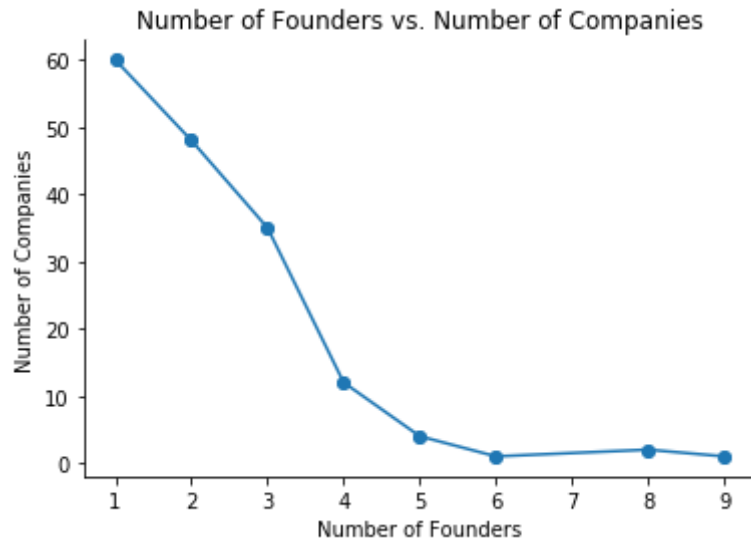
```
In [90]: fix, ax = plt.subplots()
ax.scatter(f.Number_Founders, f.Valuation,
           alpha= 0.50)
ax.set_title("Number of Founders vs. Valuation")
ax.set_ylabel("Valuation")
ax.set_xlabel("Number of Founders")
ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)
plt.plot(f.Number_Founders, f.Valuation, '-o')

plt.show()
fix, ax = plt.subplots()
ax.scatter(found['#_of_Founders'], found['#_of_Companies'],
```

```
alpha= 0.50)
ax.set_title("Number of Founders vs. Number of Companies")
ax.set_ylabel("Number of Companies")
ax.set_xlabel("Number of Founders")
ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)
plt.plot(found['#_of_Founders'], found['#_of_Companies'], '-o')

plt.show()
```





### Founders Results

Previously, we found that most startups have only 1-3 founders. However, looking at the number of founders by the average valuation, the highest are the founding teams of 2, 5, and 9. In Part 1, there were only a handful of companies who had founders with a team of 5 or 9. This suggests that what sets an above average unicorn from the rest, is having more than 1 founding member. This aligns with the Y-Combinator's style of choosing startups based on the quality of the founding team, key word "team" not individual.

## 2. Investors

I took the top 20 highest valued companies and analyzed their Investors. I used the Counter and Panda method again because there are many different investors so it would be easier than grouping them by a unique string of investors.

```
In [73]: top_uni = uni.sort_values(by = 'Valuation', ascending = False)
          .head(20)
          top_uni.head()
```

Out[73]:

	Name	Slogan	Description	Categories	Location	Founders
--	------	--------	-------------	------------	----------	----------



	Name	Slogan	Description	Categories	Location	Found
1	Uber	Everyone's Private Driver	King of ride-hailing apps, Uber now struggles ...	[Cars, Marketplace, Mobile, Unicorn]	San Francisco, CA	2009
2	Xiaomi	Only for fans	Centering the business around devoted fans, Xi...	[Hardware, Mobile, Unicorn]	Beijing	2010
3	Airbnb	Belong Anywhere	Airbnb's online platform lets users rent uniqu...	[Travel, Marketplace, Unicorn]	San Francisco, CA	2008
4	Palantir	NaN	Palantir sells analytics software for integrat...	[Big Data, Security, Analytics, Unicorn]	Palo Alto, CA	2004
5	Snapchat	NaN	Snapchat's photo and video messaging app with ...	[Messaging, Social Media, Mobile, Unicorn]	Los Angeles, CA	2011

```
In [74]: new_inv = Counter(dict(Counter(" ".join(top_uni.Investors).split(',')).most_common(20)))
```

```
In [75]: new_inv = pd.DataFrame.from_dict(new_inv, orient='index').reset_index()
```

```
In [76]: new_inv.columns= ['Investor', '# of Top 20 Unicorns Invested In']
new_inv
```

Out[76]:

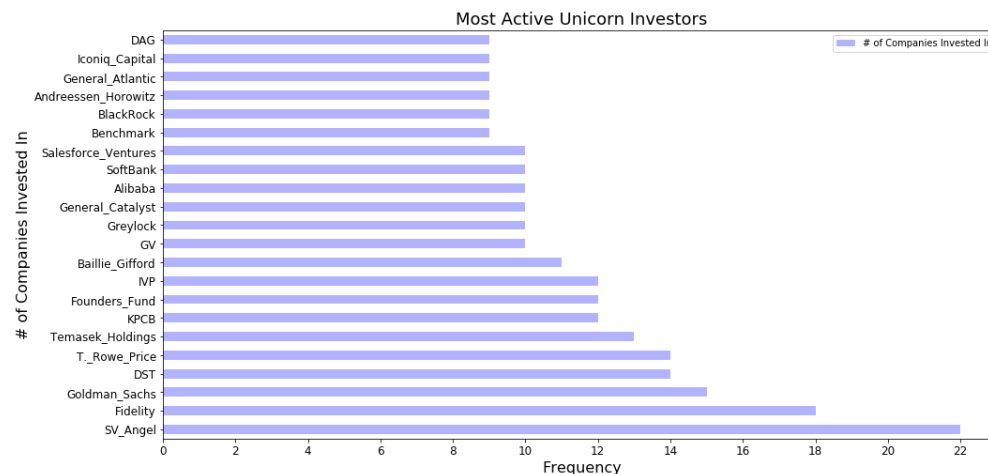
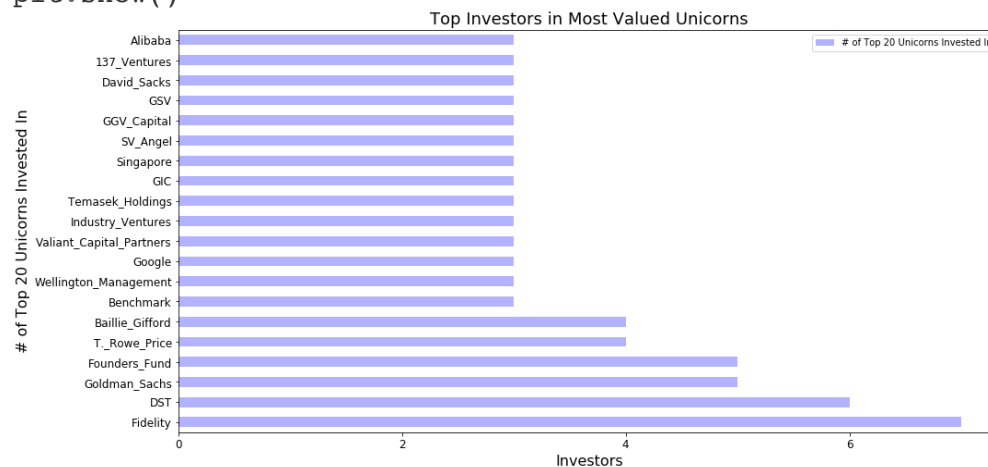
	Investor	# of Top 20 Unicorns Invested In
0	Fidelity	7
1	DST	6
2	Goldman_Sachs	5
3	Founders_Fund	5
4	T_Rowe_Price	4
5	Baillie_Gifford	4
6	Benchmark	3
7	Wellington_Management	3
8	Google	3
9	Valiant_Capital_Partners	3
10	Industry_Ventures	3
11	Temasek_Holdings	3
12	GIC	3
13	Singapore	3
14	SV_Angel	3
15	GGV_Capital	3
16	GSV	3
17	David_Sacks	3
18	137_Ventures	3
19	Alibaba	3

The data is best visualized with a bar chart.

```
In [92]: new_inv.plot(x='Investor', y='# of Top 20 Unicorns Invested
In', kind='barh', xticks = range(0,8,2),color = 'b', stacked =
False, fontsize = 12, figsize = (16,8), alpha = 0.3)
```

```
plt.xlabel('Investors', fontsize = 16)
plt.ylabel('# of Top 20 Unicorns Invested In', fontsize = 16)
plt.title('Top Investors in Most Valued Unicorns', fontsize = 18)
plt.show()

inv.plot(x='Investor', y='# of Companies Invested In', kind='barh',
xticks = range(0,23,2),color = 'b', stacked = False,
        fontsize = 12, figsize = (16,8), alpha = 0.3)
plt.xlabel('Frequency', fontsize = 16)
plt.ylabel('# of Companies Invested In', fontsize = 16)
plt.title('Most Active Unicorn Investors', fontsize = 18)
plt.show()
```



### Top Investors Results

There is little difference between the results from the other Investors analysis. There seems to be more Venture Capital firms investing in highly valued companies but it's more or less still Fidelity, Goldman Sachs, and other banks. There is more tech companies as investors like Google, and Alibaba.

## 3. Growing Industries

Again, I will be looking at the categories and the buzzwords in the description to identify major industries for the most valued unicorn companies.

### Most Valued Industry Categories

I will use Counter and Panda to record the frequency of the top 20 valued unicorns from the top\_uni dataframe.

```
In [93]: new_cat1 = Counter(dict(Counter(top_uni.Categories_001)))
new_cat2 = Counter(dict(Counter(top_uni.Categories_002)))
new_cat3 = Counter(dict(Counter(top_uni.Categories_003)))
new_cat4 = Counter(dict(Counter(top_uni.Categories_004)))
new_cat5 = Counter(dict(Counter(top_uni.Categories_005)))
```

```
In [94]: new_cat_sum = new_cat1 + new_cat2 + new_cat3 + new_cat4 + new_cat5
```

```
In [95]: new_cats = pd.DataFrame.from_dict(new_cat_sum, orient='index').reset_index()
new_cats.columns = ['Word', 'Frequency']
new_cats = new_cats.sort_values(by = ('Frequency'), ascending=False)
new_cats.head()
```

Out[95]:

	Word	Frequency
24	NaN	39
20	Unicorn	20

	Word	Frequency
18	Mobile	7
17	Marketplace	6
0	Cars	4

```
In [96]: new_cats = new_cats.drop([24], axis = 0)
```

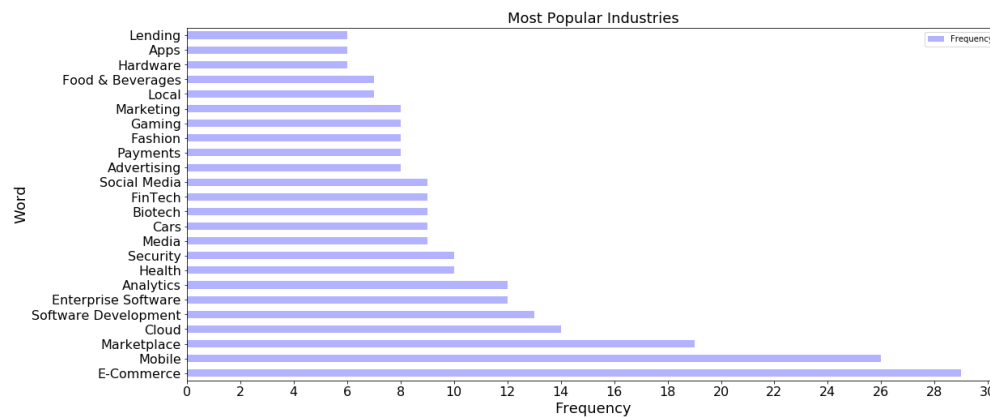
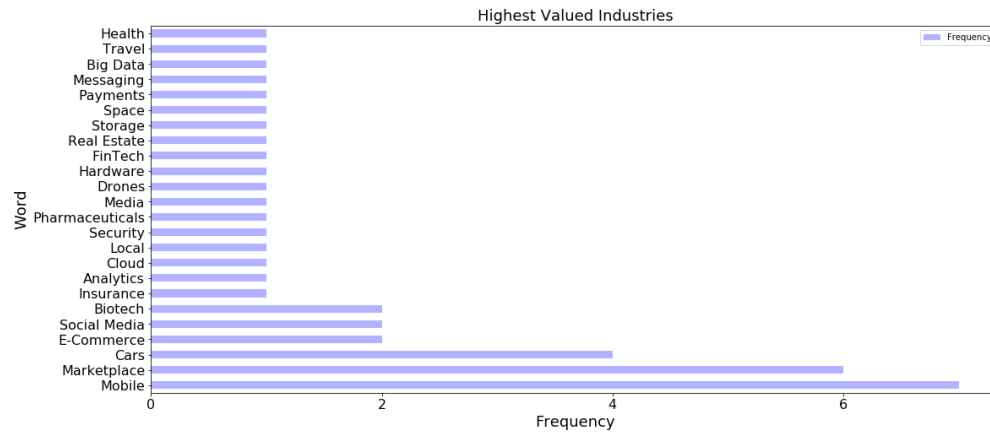
```
In [97]: new_cats = new_cats.drop([20], axis = 0)
```

```
In [99]: new_cats.head()
```

Out[99]:

	Word	Frequency
18	Mobile	7
17	Marketplace	6
0	Cars	4
5	E-Commerce	2
8	Social Media	2

```
In [100]: new_cats.plot(x='Word', y='Frequency', kind='barh', xticks =
    range(0,8,2),color = 'b', stacked = False, fontsize = 16, f
    igsiz=(18,8), alpha = 0.3)
plt.xlabel('Frequency', fontsize = 18)
plt.ylabel('Word', fontsize = 18)
plt.title('Highest Valued Industries', fontsize = 18)
plt.show()
cats.plot(x='Word', y='Frequency', kind='barh', xticks = ran
    ge(0,32,2),color = 'b', stacked = False, fontsize = 16, figs
    ize=(18,8), alpha = 0.3)
plt.xlabel('Frequency', fontsize = 18)
plt.ylabel('Word', fontsize = 18)
plt.title('Most Popular Industries', fontsize = 18)
plt.show()
```



### Industry Results

The most valued categories are still Mobile and E-Commerce like the most popular graph. But at the top of the most valued list is also cars, and social media, biotech and insurance. This makes sense because the barrier to entry for E-Commerce and Mobile is relatively low so there will be higher quantities of e-commerce companies, which means more investment activity in it. For cars, social media, biotech, and insurance, the barrier to entry is much higher. So even though they are lower on the most popular industry scale, since there is less competition for capital, they can have higher valuations. For example look at Facebook, SnapChat and Twitter, there are only a few social media platforms and it's hard to build the customer base.

### Buzzwords

I'm going to approach buzzwords by using the top\_uni frame with the top 20 values companies and use the Counter and Panda method

```
In [102]: top_uni.Description.str.rsplits(pat=' ').head()
```

```
Out[102]: 1    [King, of, ride-hailing, apps,, Uber, now, str...
2    [Centering, the, business, around, devoted, fa...
3    [Airbnb's, online, platform, lets, users, rent...
4    [Palantir, sells, analytics, software, for, in...
5    [Snapchat's, photo, and, video, messaging, app...
Name: Description, dtype: object
```

```
In [103]: new_results = set()
top_uni.Description.str.lower().str.split().apply(results.update)
print (results)
```

```
{'product', 'shipments', 'asia's', 'curevac', 'revenue', 'brokerage', 'uptake', 'security.', 'gaming.', '15%', 'brand', 'snapdeal', 'than', 'ingestible', 'sparse.', '2,000', 'return', 'request', 'buys', 'banks.', 'presenting', 'boards', 'homes', 'up', 'applied', 'ucar', 'fees.', 'things', 'communication.', 'docker's', 'select', 'chains', 'njoy', 'relatively', 'work', 'listings', 'microsoft', 'save', 'use', 'asia.', 'powa', 'seats', 'defending', 'systems', 'pioneer', 'volume', 'coupa', 'likely', 'wide', 'multiplayer', 'packages.', 'prove', 'lets', 'and', 'flexibility', 'accepts', 'america', 'scenes', 'india's', 'service', 'plans', 'containers', 'check', 'banks', 'maternity', 'airbnb's', '5%', 'appnexus's', 'credit.', 'like.', 'mortgages', 'leading', 'gas', 'market.', 'empty', 'finance', 'quality', 'top', 'client', 'media.', 'coming', 'industry', 'licensing', 'creators', 'rocket', 'draw', 'make', 'music', 'alternative.', 'beibei's', 'red', 'gilt', 'reservations', 'delivery.', 'it', 'plans', 'some', 'expensive', 'launches', 'lenders', 'policy.', 'grabtaxi', 'hotels', 'detection.', 'search', 'utility', 'shazam', 'mid-', 'developer', 'main', 'wifi', 'inspect', 'friendly', 'what', 'zeta', 'hr.', 'recent', 'retail', 'tissues', 'individual', 'performance.', 'comes', 'storing', 'smaller', 'version', 'slims', 'delivering', 'for', 'interoperability', 'commission', 'dinners', 'listings', 'firm', 'pressure', 'readers', 'dominant', 'subscribers.', 'promising', 'help', 'outlet', 'king', 'self-destructing', 'income', 'not', 'hadoop.', 'keeps', 'automatin
```

g', 'deliver', 'another', 'state', 'hours', 'responsible',  
 'go', 'territory', 'transaction-ready.', 'fraud', 'originat  
 ing', 'plus', 'adyen', 'communication', 'commitment', 'ren  
 tal.', 'health', 'zone', 'merger', "'s", 'toward', 'recipe  
 s', 'website', 'optimize', 'good', 'latin', 'runs', '(mrn  
 a)', 'immune', 'offerings', 'airbnb', 'free', 'square', 'm  
 ost', 'mobile-app', 'controlling', 'rates', 'introducing',  
 'headsets', 'supports', 'include', 'holds', 'own', 'geneti  
 c', "group's", 'software', 'lending', 'pushing', 'revenu  
 e.', 'advantages', 'sharing', 'east.', 'products.', 'transa  
 ctions', 'uers', 'systems.', 'headset', 'businesses', 'worl  
 d', 'tips', 'pays', 'tujia', 'beer', 'personal', 'uses',  
 'excellent', 'connect', 'beauty', '500', 'proxies', 'brand  
 s', 'pictures', 'subsidiaries', 'insurance', 'houzz', 'adv  
 ertising', 'known', 'human', 'focus', 'existing', 'blood',  
 'koudai', 'high-earners', 'week', 'friendlier', 'audience  
 s.', 'profits', 'onto', 'leader', 'generalized', 'includin  
 g', 'catalog', 'brands.', 'connected', 'targets', 'loyalt  
 y', 'broker', 'images', 'along', 'sidestepping', 'opportun  
 ities.', 'ad', 'restaurants.', 'insurance.', 'peripherals',  
 'services', 'gaming', 'trackers', 'store', 'see', 'power  
 s', 'does', 'zip', 'zhong', 'security', 'optimizing', 'ente  
 rprise', 'corresponding', 'employees', 'automates', 'desig  
 ned', 'servers', 'memory', 'speeds.', 'mia.com', 'benefit  
 ed', 'rna', 'opportunities', 'medallia', 'analytics.', 'pat  
 ients', 'interchange', 'continues', 'implant', 'promises',  
 'lakala', 'lifestyle', 'healthy', 'whitens.', 'child', 're  
 ferral', 'wearable', 'didi', 'apps', 'ventures', 'xiaomi',  
 'tickets', 'energy-intensive', 'carpooling', 'instacart',  
 'users.', 'manufactures', 'gambling.', 'established', "snap  
 chat's", 'are', 'buy', 'company', 'self-optimizing', 'casu  
 alty', 'except', 'source', 'decision-making', '2006', 'clie  
 nts.', 'ones.', 'deem', 'mortar', 'aiming', 'address', 'uni  
 que', 'twilio', 'high-performance', 'also', 'jet', 'huge',  
 'relying', 'zomato', 'vacation', 'vetted', 'burning', 'we  
 bsite', 'carrying', 'consumers', 'lufax', 'china's', 'capt  
 ures', 'million', 'subprime', 'this', 'become', 'parentin  
 g', 'telemedicine', 'single', 'internationally', 'celebrit  
 y', 'using', 'decolar', 'shoppers', 'features', 'illumio',  
 'visual', 'growing', 'bill', 'larger', 'programs', 'autoric  
 kshaws', 'things.', 'fanduel', '23andme', 'booking', 'lyf  
 t', 'people', 'delivers', 'bots.', 'industries.', 'inmobi',  
 'original', 'project', 'e-cigarettes', 'members', 'throug



h', 'group', 'analytic', 'models.', 'struggles', 'spendin  
g,', 'insidesales', 'online', 'controls.', 'them', 'bla',  
'faster', 'amazon', 'behind', 'vine', 'develop', 'autol',  
'variety', 'extensive', 'connections,', 'insurance', 'letti  
ng', 'device.', 'mainstay', 'computing', 'shop.', 'relies',  
'networking', 'bros.', 'designers', '(mrna', 'rapidly', 'mi  
nus', 'blocked', 'internet', 'which', 'effects', 'trendy',  
'slower', 'backup.', 'music,', 'southeast', 'restauranteer  
s.', 'firm,', 'sometimes', 'under', "mongodb's", 'struggl  
e', 'government.', 'class', 'hero', 'mostly,', 'and,', 'bri  
ck', 'wepiao', 'adware', 'developers', 'digital', 'companie  
s.', 'that', 'cloudera', 'solution', 'foodie.', 'prosper',  
"slack's", 'further', 'b2b', 'therapies', 'goods.', 'shifte  
d', 'analysis.', 'record', 'channels.', 'molecular', 'gatew  
ays', 'sprinkl', 'glam', 'anti-virus', 'countries.', 'aid  
s', 'how', 'event', 'consoles,', "mulesoft's", 'infinidat',  
"kabam's", 'discussed', 'kind', 'it', 'tech', 'software.',  
'real-time', "bloom's", 'community', 'service,', 'message  
s', 'devoted', 'kudu,', 'domo', 'taxis', 'deliveries', 'red  
undant', 'producers.', 'platform.', 'android', 'app', 'inve  
ntory', 'themselves.', 'heavyweights,', 'management.', 'fan  
li', 'young', 'outside', 'difficult-to-destroy', 'be', 'buz  
zfeed', 'hopes', 'thrives', 'centered', 'one', 'sample', 'c  
omplaints', 'proteus', 'fans,', 'dji', 'klarna', 'behalf',  
'major', 'dollars', 'body,', 'home.', 'integrates', 'adtec  
h', 'east,', 'automatically', 'eyeglasses', 'hosts', 'busin  
esses,', 'office', 'safety', 'its', 'mouse', 'sites', 'uk-b  
ased', 'spending', 'code.', 'marketplace.', 'universal', 'y  
et.', 'ticketing', 'salesforce', 'apis', 'past', 'instituti  
ons', 'seen', 'different', 'tango', 'asia,', 'primary', 'su  
ite', 'r&d', 'extremely', 'by', 'questionnaire', 'integrati  
ng,', 'electricity', 'beer', 'media,', 'without', 'china.',  
'card', 'boutiques,', 'like', 'mogujie', 'well', 'ship', 's  
cale', 'wine.', 'china', 'integration.', 'returns.', 'suc  
h', 'launcher,', 'technical', '20', '100', 'easy-to-use',  
'would', 'sales', 'buyer,', 'professionals', 'stem', 'lend  
s', 'relationship', 'many', 'feedback,', 'commission.', 'ac  
cept', 'difficult', 'technology', 'members-only', 'appdirec  
t', 'advanced', 'converts', 'set.', 'diabetes', 'centerin  
g', 'revenues', 'office.', 'autonomous', 'exchange', 'contr  
ol,', 'marklogic', 'regulation.', 'yelp-like', 'electronic  
s', 'personalized', 'particularly', 'taobao', 'drawing', 'a  
ttacks,', 'designs', 'chat.', 'vendor', 'listings.', 'deale

rships.', 'scene,', 'demand', 'providing', 'meitu', 'recent  
 ly', 'need', 'messaging,', 'advertising,', 'of', 'draws',  
 'marketing,', 'clients', 'extra', 'jasper', 'efficient,',  
 'data.', 'zscaler', 'blog,', 'discovery,', 'business,', 'sa  
 vings.', 'treatments', 'wine', 'transactions,', '"they're",  
 'revolutionized', 'platform', 'passes', 'data-driven', 'fea  
 tures,', 'factors', 'offers', 'note-taking', 'month.', 'des  
 igner', 'bundled', 'spanning', 'improvement,', 'operates',  
 'accomodations', 'wi-fi', 'efficient', 'lazada', 'dna', 'co  
 mpeting', 'car.', 'shared', 'passing', 'projects', 'cloudfl  
 are', 'fees,', 'apttus', 'lookout', 'video-driven', 'youn  
 g.', 'environments.', 'free-to-play', 'around', 'per', 'ref  
 erred', 'price', 'chinese', 'uber.', 'well-recognized', 'ty  
 pically', 'margin,', 'fashion-oriented', 'movie', 'event  
 s,', 'cuztomizable', 'apron', 'ad-supported', 'benefits',  
 'company', 'hotspots', 'add', 'wireless', 'spacecraft,', 'e  
 nvironment', 'parts', 'focusing', 'estate', 'moderna', 'i  
 s', 'cities', 'startups', 'scan', 'high', 'shop', 'number',  
 'offer', 'payments,', 'salaries', 'centers,', 'tencent',  
 'releases', 'publishers.', 'killing', 'nutritious', 'job',  
 'expanding', 'physical', 'respect.', 'cars', 'designs,', 'c  
 ompanies,', 'jiuxian', 'mover', 'innovations', 'emerging',  
 "israel's", 'mathematics', 'scanning', 'all', 'point', "ada  
 ptive's", 'retailer', 'emphasis', 'nutritious,', 'money',  
 'optimization,', 'mobile,', 'security,', 'manages', 'can',  
 'contract', 'trackers.', 'an', 'sports', 'them.', 'gives',  
 'printing', 'avant', 'cheaper.', 'several', 'markups', 'gla  
 ssdoor', 'cells.', 'quick,', 'industry', 'filters.', 'crea  
 tes', 'fees', 'marketing', 'thumbtack', 'in', 'announced',  
 'operating', 'improve', 'generation', 'wework's', 'more',  
 'now', 'palantir', "shazam's", 'format', 'both', 'course',  
 'loans', 'close', 'within', 'experience.', 'there', 'apach  
 e', 'calls', 'results.', 'video', 'hidden', 'generates', 's  
 urvey', 'smartphone', 'marketplac', 'business.', 'docusig  
 n', 'ancestry.', 'unity.', 'access', 'open-source', 'inform  
 ation.', 'us,', 'fantasy', 'monetizes', 'appdyanmics', 'len  
 ding,', 'experience', 'rental', 'second', 'diagnostics', 'o  
 thers', 'classy', 'quikr', 'phones,', 'voice', 'craigslis  
 t', 'is', 'those', 'deals', 'accessories', 'down.', 'oxfor  
 d', 'space', 'spacex', 'photo-sharing', 'blue', 'against',  
 'populations', 'apus', 'russia', 'post', 'items.', 'conten  
 t', 'backup', '1,000', 'dropbox', 'infant-care', 'lender',  
 'registered', 'series', 'regulatory', 'cruises,', 'dependin

g', 'sites,', 'enterprise,', 'learning', 'machines.', 'side', 'relax', 'shopping,', 'rebates.', 'chattiness', 'peak', 'neighborhoods.', 'developer', 'chat', 'tool,', 'jawbone', 'databases', 'advertising.', 'list', 'legally', 'loans,', 'reducing', 'sale', 'board', 'these', 'file', 'exchange,', 'clothes', 'unlimited', 'logins', 'crowdsourcing', 'marketplace', 'systems,', '12', 'airbnb', 'prosper', 'cancer.', 'distributes', 'companies', 'developer-friendly', 'authentic', 'purchase', 'mrna', 'rich', 'potentially', 'being', 'imports,', 'players.', 'your', 'ever,', 'simplivity', 'opportunity', 'yielding', 'only', 'udacity', 'expenses', 'stores', 'low', 'firms,', 'free,', 'throughout', 'low-end', 'applications,', 'food', 'deal', 'clean', 'combat,', '70+', 'surveys,', 'result.', 'loans.', 'usable', 'devices.', 'available', 'infrastructure', '2011,', 'blocking', 'devices', 'reasonable', 'phones', 'south', 'deals.', 'approach', 'been', 'raised', 'hit', 'done', 'drivers', 'benefit', 'bloggers', 'targeting', 'allows', 'machine', 'pictures.', "it's", 'has', 'regulations,', 'lost', 'appstore', 'ever', 'do', 'order', 'happiness.', 'year', 'licensed', 'api', 'data,', 'serious', 'entity', 'smooth', 'real', 'alternative', 'swipes,', 'android,', 'when', 'rate.', 'shopping', 'immuno-oncology', 'mobile', 'ping', 'protein', 'fortune', 'nutanix', 'games.', 'cost.', 'accuracy', 'malicious', 'records.', 'big', 'wearables.', 'markets', 'foreign', 'specializes', 'easier', 'multiple', 'reuse', 'theaters.', 'much', 'developers,', 'banking', 'rather', 'reach', 'to', 'send', 'anypoint', 'attitude', 'options', 'gimbal', 'adds', 'employee', 'checks', 'tools', 'concentrated', 'support,', 'documents', 'windows,', 'connects', 'otherwise.', 'traffic,', 'secure', 'eyes', 'e-commerce', 'premium', 'end', 'fun', 'tools.', 'hardware', 'maker,', 'kabbage', 'adults.', '2008', 'orders', 'satire', 'surgery', 'provides', 'global', 'borrowers.', 'augmented', 'datto', 'fee,', 'publishing,', "spotify's", 'gaming', 'merchant,', 'putting', 'long', 'warehousing', 'come', 'before.', 'selling', 'management', 'taking', 'products', 'customers', 'where', 'innovative', 'vintage-inspired', 'try', 'win', 'vehicle', 'signals', 'signing', 'earliest', 'gusto', 'service.', 'intelligence', 'cancers.', '70', 'transmission,', 'a', 'sign-in.', 'investors,', 'bookmarking', 'avast', 'interface', 'few', 'drugs.', 'apps', 'information', 'freemium', 'monitors', 'authenticity', 'aggressively', 'reports.', 'drugs,', 'offering', 'email', 'cross

-device', 'eight', 'training', 'photography', 'wall', 'aiwu  
jiwu', 'consumer', 'app,', 'company's', 'customer', 'draftk  
ings', 'cells,', 'form', 'menus', 'surveymonkey', 'user',  
'mothers', 'from', 'abdominal', 'kinds,', 'enlist', 'cell',  
'forms', 'elite', 'sends', 'messaging', 'storage.', 'warne  
r', 'sensors', 'every', 'merchants.', 'china's', 'different  
iate', 'law', 'create', 'alliance', 'video,', 'vice', 'trea  
ts', 'fourth', 'hundreds', 'delivery,', 'uncommon,', 'insti  
tution', 'search,', 'defend', 'increasing', 'cloud-based',  
'across', 'free.', 'monitoring', 'large', 'buying', 'remot  
e', 'reverse', 'developing', 'works.', 'transferwise', 'gre  
ener', 'hollywood,', 'helps', 'checkout', 'hellofresh', 'co  
re', 'fastest', 'massively', 'encrypting', 'alibaba,', 'mes  
senger', 'doctor's', 'opportunity,', 'computers.', 'cance  
r', 'as', 'open', 'retention.', 'explain', 'medicine', 'se  
o,', 'miaopai', 'organized', 'but', 'lamabang', 'created',  
'cut', 'generation,', 'labels', 'cyberattacks.', 'include  
s', 'purchases.', 'ride-hailing', 'password', 'over', 'gith  
ub', 'with', 'credit', 'previously', 'an.', 'tracking', 'de  
stination', 'works', 'in-game', 'intarcia', 'support', 'mas  
sive', 'any', 'qualtrics', 'programs.', 'hadoop', 'early',  
'based', 'content.', 'evernote', 'strengthen', 'riders', 'p  
rimarily', 'communications,', 'risk-management', "organizat  
ion's", 'analyzes', 'microchip', 'moving', 'merchants', 'se  
quencing', 'no', 'vetted', 'media', 'payments', 'medical',  
'chuxing', 'e', 'shops.', 'interactive', 'startup', 'app  
s.', 'request,', 'rate', 'imported', 'capacity', 'ticket',  
'stolen', '30', 'mode', 'high-end', 'year.', '5', 'gourme  
t', 'key', "amazon's", 'employees', 'pinterest', 'improve  
d', 'simplifies', 'allow', 'starting', 'blablabla.', 'logis  
tics', 'lower-mid', 'visualizing,', 'prices.', 'mess', 'cal  
culation.', 'ads', 'avoiding', 'player', 'recovery.', 'smal  
l', 'spotty.', 'vox', 'less', 'about', 'site,', 'applying',  
'rent', 'produces', 'lifestyle,', 'private', 'guaranteein  
g', 'co-produced', 'zenefits', 'high-capacity,', 'other',  
'leap', 'development,', 'server', 'art', 'advice,', 'hailin  
g', 'analyze', 'interpersonal', 'application', 'laws', 'bec  
ause', 'oscar', 'conscious', 'credentials', 'targeted', 'e  
n', 'kik', 'medium', 'let', 'cash-on-delivery', 'dianpin  
g,', 'mongodb.', 'track', 'technologies,', 'alba', 'educati  
on', 'architects', 'optimized', 'sources', 'payments.', 'pr  
omise', 'enterprise.', 'referrals.', 'karma', 'monthly', "m  
ozido's", 'two.', 'still', 'returns', 'screening', 'voic

e,', 'total', 'daily', 'razer', 'reach.', 'india,', 'son  
g.', 'reports', 'change', 'streaming', 'sellers.', 'copy',  
'wechat,', 'rewards', 'market', 'european', 'bank', 'provid  
er,', 'all-in-one', 'minimalist', 'game', 'take', 'lower',  
'market,', 'investors', 'sysadmins', 'carriers,', 'classroo  
m', 'bookings,', 'property', 'nosql', 'furniture.', 'lead  
s.', 'borrower', 'chauffeured', 'thousands', 'sales,', 'ube  
r', 'sms,', 'flight', 'portable', 'it.', 'cash', 'eventbrit  
e', 'roadtrip', 'services', 'attracting', '3d', 'code,', 'f  
ocused', 'student', 'age', 'methods', 'industrial', 'develo  
ps', 'prices,', 'ubiquitous', 'discount', 'meituan.com,',  
'drives.', 'theranos', 'workers', 'lenders.', 'focuses', 'b  
io-engineering', 'phone', 'when.', 'lead', 'dining.', 'univ  
ersity,', 'while', 'analyzing', 'takes', 'range', 'everythi  
ng', 'better', 'volume.', 'retaining', 'photo', "firm's",  
'you', 'originally', 'network,', 'websites', 'borrowers',  
'itself', 'social', 'trading', 'hyperconverged', 'commerc  
e', 'ingredients', 'expanded', 'then', 'pluralsight', 'shar  
e', 'the', 'processes', 'have', 'separate', 'p2p', 'pins',  
'publishers,', 'into', 'homemaking,', 'reality', 'databas  
e,', 'drone', 'billion', 'storage', 'users', 'provider', 'r  
efinance', 'large-scale', 'wordpress,', 'ground', 'larges  
t', 'grocery', 'beats', 'mostly', 'facebook', 'panshi', 'ca  
tegories', 'retailer,', 'claims', 'nextdoor', '18-34', 'cmb  
ines', 'novel', 'energy', 'helping', 'tracks', 'networks.',  
'network', 'mall', 'options.', 'welcome', 'courses', 'polym  
er', 'requiring', 'cloud', 'tests,', 'lines,', 'goes', 'poi  
nt-of-sale', 'online-only', 'augment', 'retouch', 'contro  
l', "customers'", 'reviews', 'recommendations', 'document',  
'magic', 'retailers', 'banning', 'point,', 'forum', 'paymen  
t', 'each', '10', 'journalism', 'handle', 'institutionaliz  
e', 'home,', 'pin', 'films', 'sell', 'hr', "actifio's", 'di  
gitally', 'their', 'international', 'first-mover', 'jessic  
a', 'makes', 'back-end', 'two', 'australia,', 'techniques',  
'prices', 'india.', 'brash,', 'despite', 'on', 'charging',  
'performance', 'so', 'analytics', 'valuable,', 'receding',  
'turnkey', 'viral', 'system,', 'indians', 'or', 'submit',  
'local', 'sign', 'videos,', 'easier,', 'smartphones,', 'dep  
artment', 'street', 'valuable', 'allowing', 'accounts.', 's  
cience', 'farfetch', 'song', 'brand', 'who', 'strength', 'b  
eginning', 'individuals', 'us', 'command', 'mobile.', 'solu  
tions', 'europe,', 'mac.', 'policy,', 'grown', 'employing',  
'flipkart', 'meizu', 'reserved', 'anti-terrorism', 'fashio

```
n', 'home', 'well-established', 'students', 'making', 'kore
a', 'popular', 'reviews,', 'system', 'legendary', 'masse,',
'biological', '190+', 'military', 'soon.', 'one97', 'item
s', 'monetization', 'a,', 'prepare', 'insights', 'satellite
s.', 'amasses', 'identifies', 'delivery', 'china,', 'highe
r', 'agency', 'travel', 'ironsource', 'flights,', 'peopl
e,', 'middle', 'immunizations.', 'fitness', 'officially',
'businesses.', 'speakers.', 'buyers', 'tutor', 'first', 'fa
vor', 'platforms,', 'payroll', 'became', "users'", 'distrib
ution', 'maintain', 'productivity', 'new', 'move,', 'fee',
'liquor,', 'installers,', 'billing', 'premier', 'among', 'q
uickly', 'car', 'english', 'pills', 'giving', 'parents,',
'between', 'female', 'wish', 'manufactures,', 'bid', 'shoul
d.', 'seconds,', 'customers,', 'others.', 'restaurant', 'ge
ts', 'handles', 'technology,', 'may', 'at', 'use', 'games',
'consumer,', 'concepts', 'run', 'brazil,', 'skeleton', 'co-
financed', 'qualification', 'cars,', 'natural', 'sized', 'r
ates.', 'keep', 'they', "tanium's", 'records', 'data', 'use
d', 'running', 'staff', 'reporting,', 'successful', 'faste
r,', 'vendors', 'internet,', 'therapy', 'kabam', 'warehouse
s,', 'prescription', 'okta', 'taxi', 'justfab', 'garena',
'foot', 'stories', 'expense', 'makers', 'collaboration.',
'automated', 'business', 'site', 'groceries', 'collects',
'rockets,', 'times', 'bloodwork', 'charges', 'doctors,', 'o
rdering,', 'suppliers', 'compete', 'concessions', 'browse',
'appropriate', 'specialized', 'merchandise.', 'sells', 'off
ered'}
```

```
In [104]: Counter(" ".join(top_uni.Description).split()).most_common(2
0)
```

```
Out[104]: [('and', 23),
('a', 15),
('for', 14),
('of', 10),
('to', 10),
('the', 8),
('in', 5),
('local', 4),
('with', 4),
('A', 4),
('is', 4),
('has', 4),
```

```
( 'Uber', 3),
( 'from', 3),
( 'app', 3),
( 'on', 3),
( 'King', 2),
( 'other', 2),
( 'business', 2),
( 'designs,', 2)]
```

```
In [105]: new_buzzwords = Counter(" ".join(top_uni.Description).lower
().split())
for word in list(new_buzzwords):
    if word in ignore:
        del new_buzzwords[word]
print (new_buzzwords)
```

```
Counter({'king': 3, 'uber': 3, 'app': 3, 'other': 2, 'busin
ess': 2, 'designs,': 2, 'manufactures,': 2, 'consumer': 2,
'online': 2, 'users': 2, 'two': 2, 'video': 2, 'million':
2, 'per': 2, 'merger': 2, 'china's': 2, 'hailing': 2, 'thin
gs': 2, 'include': 2, 'dropbox': 2, 'new': 2, 'services':
2, 'insurance': 2, 'systems,': 2, 'chinese': 2, 'year.': 2,
'ride-hailing': 1, 'apps,': 1, 'struggles': 1, 'change': 1,
'regulations,': 1, 'prepare': 1, 'autonomous': 1, 'cars,':
1, 'develop': 1, 'logistics': 1, 'systems': 1, 'industrie
s.': 1, 'centering': 1, 'around': 1, 'devoted': 1, 'fans,':
1, 'xiaomi': 1, 'distributes': 1, 'phones': 1, 'electronic
s': 1, 'lower-mid': 1, 'market': 1, 'asia.': 1, 'airbnb's':
1, 'platform': 1, 'rent': 1, 'unique': 1, 'accomodations':
1, 'hosts': 1, '190+': 1, 'countries.': 1, 'palantir': 1,
'sells': 1, 'analytics': 1, 'software': 1, 'integrating,':
1, 'visualizing,': 1, 'analyzing': 1, 'information.': 1, 'p
roduces': 1, 'product': 1, 'lines,': 1, 'anti-terrorism':
1, 'fraud': 1, 'detection.': 1, "snapchat's": 1, 'photo':
1, 'messaging': 1, 'self-destructing': 1, 'messages': 1, 'b
ecame': 1, 'hit': 1, 'young': 1, 'people,': 1, 'leading':
1, 'payments': 1, 'media.': 1, 'amazon': 1, 'india.': 1, 'i
ndia's': 1, 'dominant': 1, 'shopping': 1, 'destination': 1,
'20': 1, 'products': 1, 'across': 1, '70+': 1, 'categorie
s': 1, '5': 1, 'shipments': 1, 'month.': 1, 'recent': 1, 'm
eituan.com,': 1, 'daily': 1, 'deals': 1, 'site,': 1, 'dianp
ing,': 1, 'yelp-like': 1, 'business,': 1, 'china': 1, 'inte
rnet': 1, 'plus': 1, 'alliance': 1, 'strengthen': 1, 'bot
```

```

h': 1, 'businesses.': 1, 'taxi': 1, 'companies.': 1, 'did
i': 1, 'chuxing': 1, 'taxis': 1, 'rather': 1, 'than': 1, 'p
rivate': 1, 'cars': 1, 'allows': 1, 'extra': 1, 'tips': 1,
'faster': 1, 'service.': 1, 'spacex': 1, 'launches': 1, 'ad
vanced': 1, 'rockets.': 1, 'spacecraft.': 1, 'satellites.':
1, 'social': 1, 'visual': 1, 'bookmarking': 1, 'tool.': 1,
'people': 1, 'pin': 1, 'images': 1, 'boards': 1, 'save': 1,
'pictures': 1, 'they': 1, 'like.': 1, 'some': 1, 'pins': 1,
'buying': 1, 'opportunity.': 1, 'yielding': 1, 'pinterest':
1, 'revenue.': 1, 'well-recognized': 1, 'first-mover': 1,
'cloud': 1, 'storage.': 1, 'competing': 1, 'industry': 1,
'heavyweights.': 1, 'been': 1, 'introducing': 1, 'tools':
1, 'photo-sharing': 1, 'security.': 1, 'wework's': 1, 'shar
ed': 1, 'office': 1, 'space': 1, 'startups': 1, 'businesse
s': 1, 'are': 1, 'expensive': 1, 'square': 1, 'foot': 1, 'b
ut': 1, 'valuable': 1, 'networking': 1, 'opportunities.':
1, 'one': 1, "china's": 1, 'largest': 1, 'p2p': 1, 'lender
s.': 1, 'lufax': 1, 'support': 1, 'major': 1, 'targets': 1,
'borrowers.': 1, 'theranos': 1, 'promises': 1, 'technolog
y': 1, 'allowing': 1, 'smaller': 1, 'blood': 1, 'draws': 1,
'medical': 1, 'tests.': 1, 'making': 1, 'bloodwork': 1, 'ea
sier.': 1, 'faster.': 1, 'cheaper.': 1, 'music': 1, 'stream
ing': 1, 'internationally.': 1, "spotify's": 1, 'freemium':
1, 'service': 1, 'gives': 1, 'access': 1, 'huge': 1, 'catal
og': 1, 'music.': 1, 'passing': 1, 'cut': 1, 'revenues': 1,
'labels': 1, 'publishers.': 1, 'dji': 1, 'high-end': 1, 'dr
one': 1, 'technology': 1, 'including': 1, 'flight': 1, 'co
ntrol': 1, 'photography': 1, 'platforms.': 1, 'gimbal': 1,
'transmission.': 1, 'remote': 1, 'controls.': 1, 'first':
1, 'online-only': 1, 'property': 1, 'casualty': 1, 'provide
r.': 1, 'zhong': 1, 'big': 1, 'data': 1, 'automated': 1, 's
olutions': 1, 'insurance.': 1, 'created': 1, 'alibaba.': 1,
'tencent.': 1, 'ping': 1, 'an.': 1, 'established': 1, 'sepa
rate': 1, 'entity': 1, 'win': 1, 'favor': 1, 'reach': 1, 'm
arket.': 1, 'move.': 1, 'raised': 1, 'over': 1, 'billion':
1, 'dollars': 1, 'under': 1, 'successful': 1, 'car': 1, 'u
s.': 1, 'lyft': 1, 'become': 1, 'difficult': 1, 'differenti
ate': 1, 'much': 1, 'larger': 1, 'except': 1, 'receding':
1, 'commitment': 1, 'friendlier': 1, 'customer': 1, 'experi
ence.': 1, 'intarcia': 1, 'bio-engineering': 1, 'abdomina
l': 1, 'implant': 1, 'releases': 1, 'diabetes': 1, 'medicin
e': 1, 'automatically': 1, 'up': 1})

```



```
In [106]: new_buzz = pd.DataFrame.from_dict(new_buzzwords, orient='index').reset_index()
new_buzz.columns = ['Word', 'Frequency']
new_buzz.head()
```

Out[106]:

	Word	Frequency
0	king	3
1	ride-hailing	1
2	apps,	1
3	uber	3
4	struggles	1

```
In [107]: new_buzz = new_buzz.sort_values(by = ('Frequency'), ascending=False).head(16)
new_buzz = new_buzz.reset_index()
new_buzz.head()
```

Out[107]:

	index	Word	Frequency
0	0	king	3
1	3	uber	3
2	59	app	3
3	221	systems,	2
4	246	chinese	2

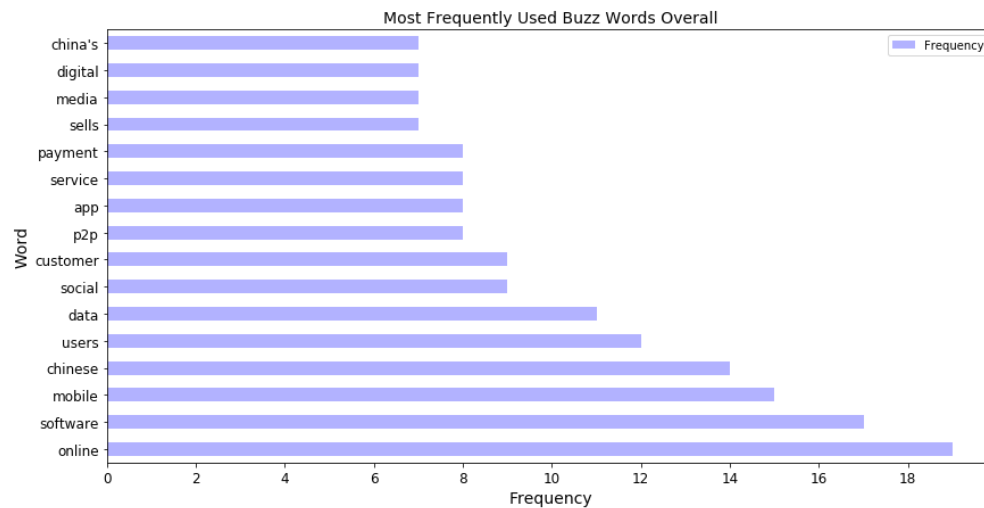
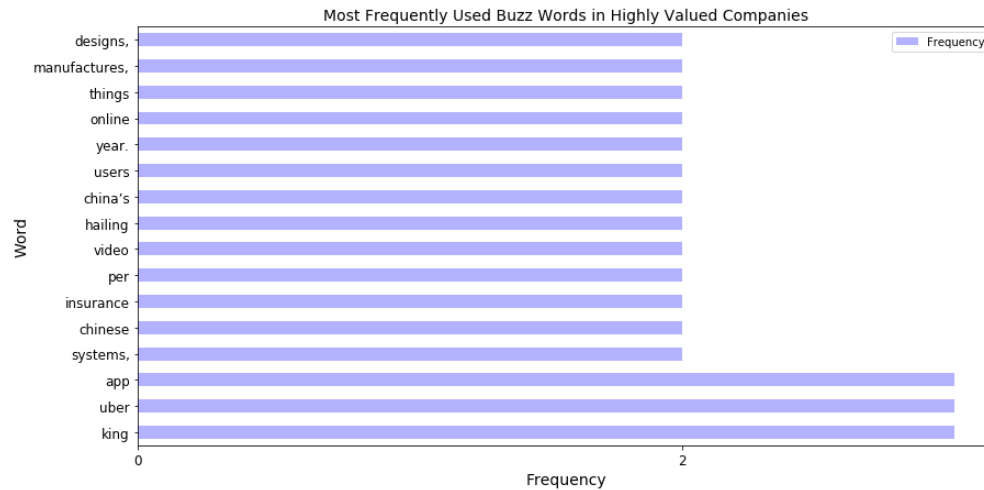
```
In [108]: new_buzz
```

Out[108]:

	index	Word	Frequency
0	0	king	3
1	3	uber	3
2	59	app	3

	index	Word	Frequency
3	221	systems,	2
4	246	chinese	2
5	181	insurance	2
6	83	per	2
7	57	video	2
8	103	hailing	2
9	101	china's	2
10	33	users	2
11	258	year.	2
12	31	online	2
13	133	things	2
14	22	manufactures,	2
15	21	designs,	2

```
In [109]: new_buzz.plot(x='Word', y='Frequency', kind='barh', xticks =
            range(0,4,2),color = 'b', stacked = False, fontsize = 12, f
            igsiz = (14,7), alpha = 0.3)
            plt.xlabel('Frequency', fontsize = 14)
            plt.ylabel('Word', fontsize = 14)
            plt.title('Most Frequently Used Buzz Words in Highly Valued
            Companies', fontsize = 14)
            plt.show()
            buzz.plot(x='Word', y='Frequency', kind='barh', xticks = ran
            ge(0,20,2),color = 'b', stacked = False, fontsize = 12, figs
            ize = (14,7), alpha = 0.3)
            plt.xlabel('Frequency', fontsize = 14)
            plt.ylabel('Word', fontsize = 14)
            plt.title('Most Frequently Used Buzz Words Overall', fontsiz
            e = 14)
            plt.show()
```



### Buzzwords Results

There are some common words between the two graphs, "China", "app", "users", and 'online'. The results from the most valuable companies were less diverse and generic, which makes sense because it is a smaller sample size as well. Since the buzzwords in the most valued companies are more diverse, I assume it is because the companies are more unique which is why they have the highest valuations.

### Conclusion

So to answer: What will the next billion-dollar start-up look like?

- *Company Profile* If it's a U.S. company it will probably be in San Francisco/ Greater San Francisco or New York. If it's in Asia it will probably be in Beijing, but if it's not in U.S. or Asia it's harder to predict. It will probably have 1 to 3 Founders.
- *Investors* The investors will most likely be venture capitalists, investment banks, or institutional investors.
- *Growing Industries* The most common industry is e-commerce, mobile and Big Data.

But to take it even further, the next Super Unicorn will probably have the following characteristics:

- *Company Profile* If it's in the super unicorn will probably be in SF or Los Angeles. If it's in Asia it will probably be in Beijing. But like a normal unicorn if it's outside Asia or the U.S., it is hard to predict. It's more likely to have a bigger team of at least 2+, with some teams as big as 9.
- *Investors* The investors are also VCs, investment banks and institutional investors but you see more big tech firms like Google and Alibaba becoming active investors.
- *Growing Industries* The categories and buzzwords are more unique, which makes sense because the above average unicorn valuation must equate an above average product or company.

Overall, as unicorn startups have become more common and valuations are skyrocketing, hopefully we can use this analysis to predict the next one.