Mining Companies' Facebook Check-In (and Other) Data

Contents

Introduction

Companies are increasingly looking to social media platforms like Facebook to collect data about (potential) customers' attitudes and engagement. For example, Facebook has a feature that allows users to "check in" to various establishment/business locations. Data about these check-ins are available for analysis.

Check-ins are (arguably) not just a proxy for counting the number of people who visit establishments, they are also a form of advertising to the customer's friends on Facebook. Therefore this metric can be a strong performance indicator for the business and may help in making performance predictions (e.g. earnings or stock prices).

In this project we will analyze a dataset from Thinknum about Facebook check-ins, likes, and 'talking-about-count' for a large group of companies. There are several questions one can consider. For example:

- Which companies have been the most successful at getting check-ins on Facebook, and do they have anything in common.
- What strategies have companies adopted to increase check-ins (if any) and do they work.
- How are check-ins, likes, and talking about count related to one another. Are they correlated/predictive of one another.
- To what extent does positive/negative news about the company or do marketing campaigns affect the metrics?
- Can these metrics be used to make useful predictions about revenue/stock prices.

Exploratory Data Analysis

The dataset is available at https://blog.thedataincubator.com/tag/data-sources/ in the "Facebook Followers" section. In this section we import the data and do a very preliminary analysis. There are about 3.6 million rows and 11 columns. The data ranges from 2015 to 2018. There are about 5000 unique "usernames", referring to company names on Facebook. For each username and most days between 2015 and 2018, the number of check-ins, 'likes', and 'talking_about_cout' are reported.

```
% imported csv as 'fb'
size(fb)
```
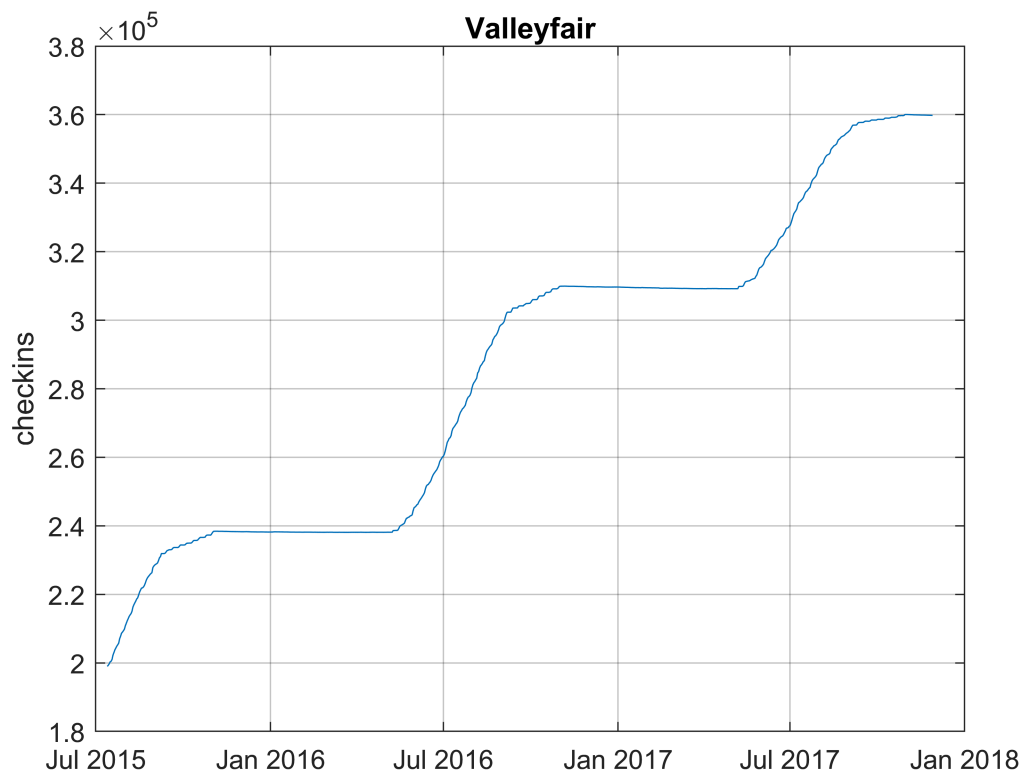
```
ans = 1×2
    3621391          11
```

Let us look at two examples:

```
cindex = fb.username == 'Disneyland';
plot(fb.time(cindex),fb.checkins(cindex))
title('Disneyland')
ylabel('checkins')
```
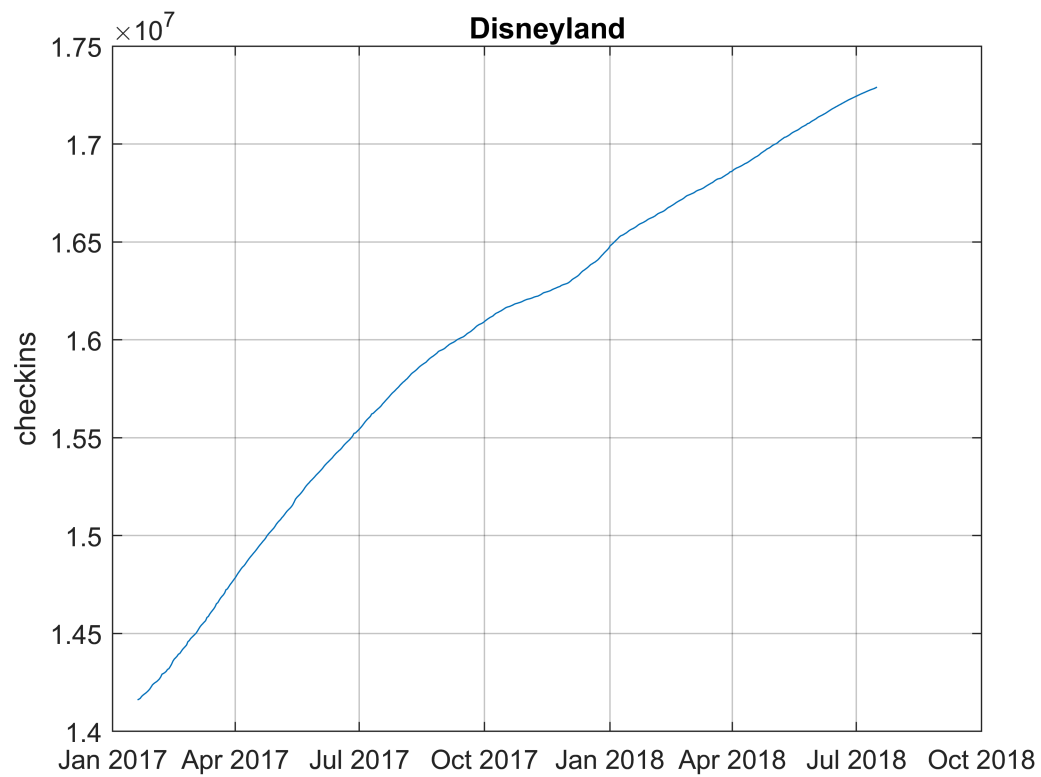
The Valleyfair amusement park seems to be closed in the winter:

```
cindex = fb.username == 'valleyfair';
plot(fb.time(cindex),fb.checkins(cindex))
title('Valleyfair')
ylabel('checkins')
```

Valleyfair

Which businesses get the most check-ins?

```
undefined = isundefined(fb.username);
fb_def = fb(~undefined,:);
companies = unique(fb_def.username);
checkins_c = [];
for c = companies'
    cindex = fb.username == c;
    checkins_c(end+1) = max(fb.checkins(cindex));
end
checkins_c = checkins_c';
```

```
company_max_checkins = table(companies,checkins_c);
most = sortrows(company_max_checkins,2,"descend");
top10 = most(1:10,:)
```

top10 = 10×2 table

	companies	checkins_c
1	Disneyland	17290550
2	WaltDisneyW...	6547775
3	DisneylandP...	4487255
4	disneylandp...	4063748
5	TheGarden	3651157

	companies	checkins_c
6	MadisonSqua...	3282939
7	empirestate...	2960168
8	KnottsBerry...	2947001
9	SeaWorld	2916920
10	BuschGarden...	2745204

```
labels = top10.companies;
temp = cellstr(labels);
ctemp = categorical(temp);
labels = reordercats(ctemp,temp);
```

```
bar(labels,top10.checkins_c)
title('Companies with the most checkins (top 10)')
ylabel('checkins')
```
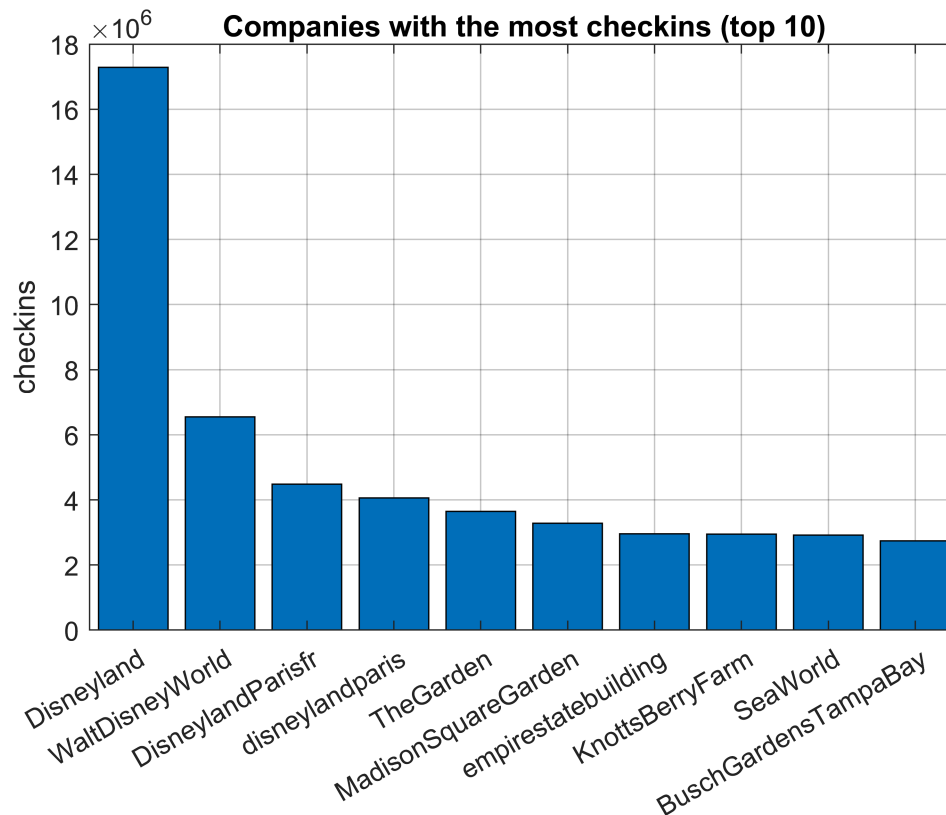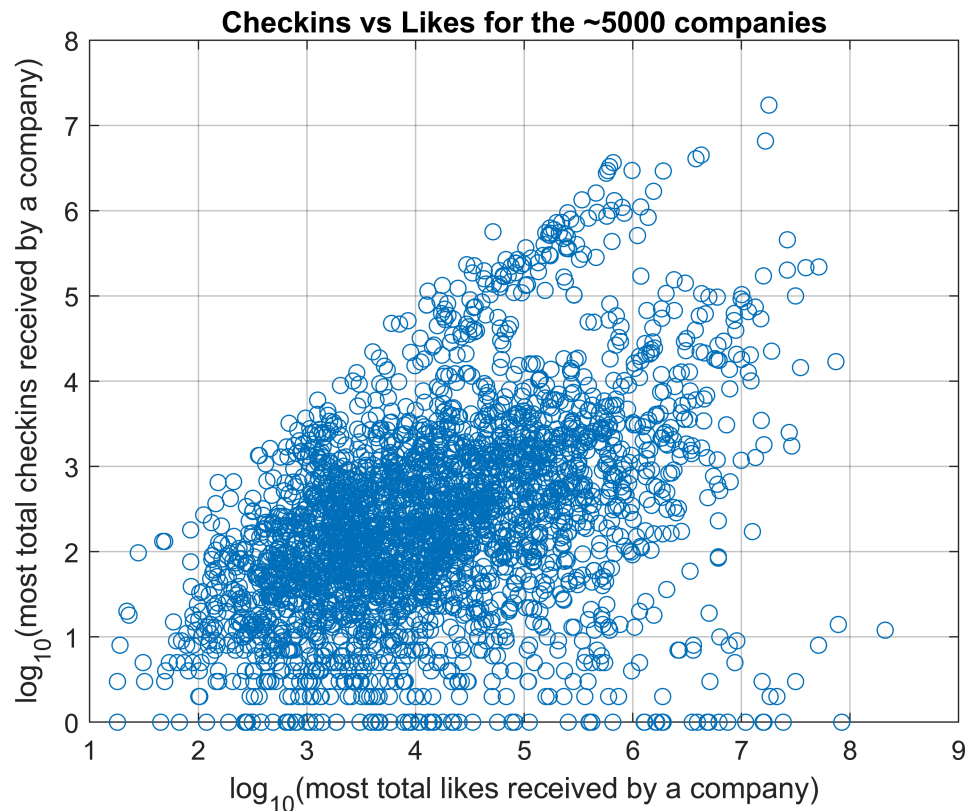

Likes vs. Check-ins

```
companies = unique(fb_def.username);
likes = [];
for c = companies'
    cindex = fb.username == c;
    likes(end+1) = max(fb.likes(cindex));
```

```
end
likes = likes'
```

```
plot(log10(likes), log10(checkins_c),'o')
title('Checkins vs Likes for the ~5000 companies')
xlabel('log_{10}(most total likes received by a company)')
ylabel('log_{10}(most total checkins received by a company)')
```
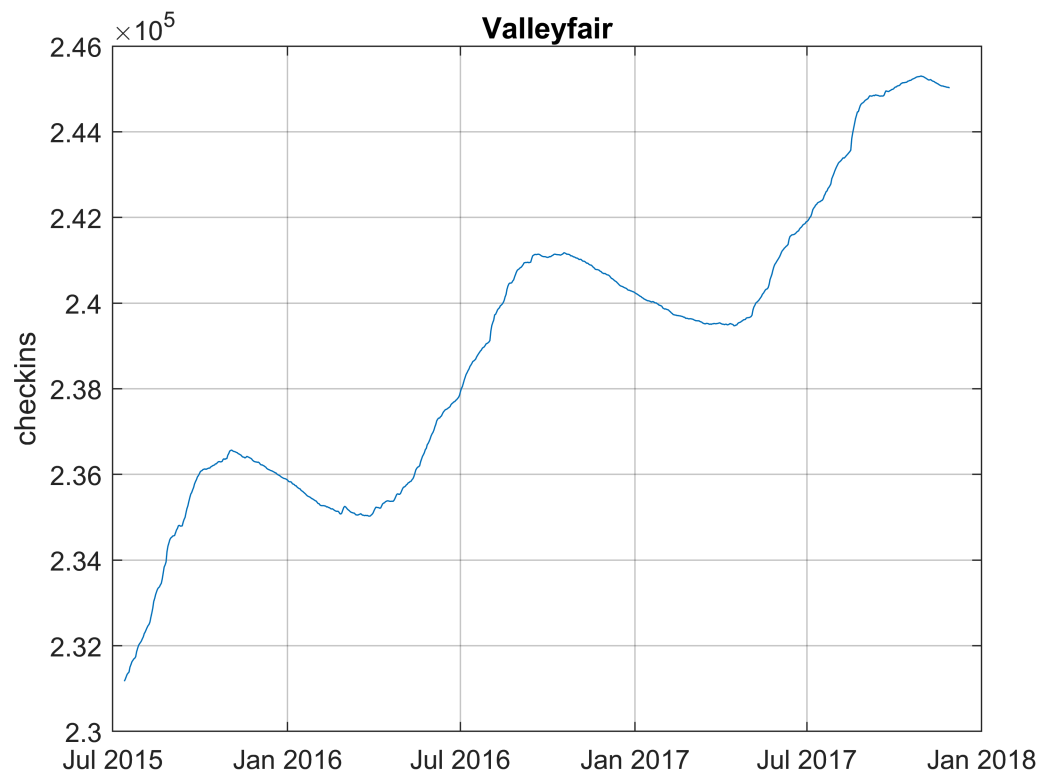

Disneyland and Valleyfair like data over time

```
cindex = fb.username == 'Disneyland';
plot(fb.time(cindex),fb.likes(cindex))
title('Disneyland')
ylabel('likes')
```

Disneyland

likes vs. time (Jan 2017 – Oct 2018)
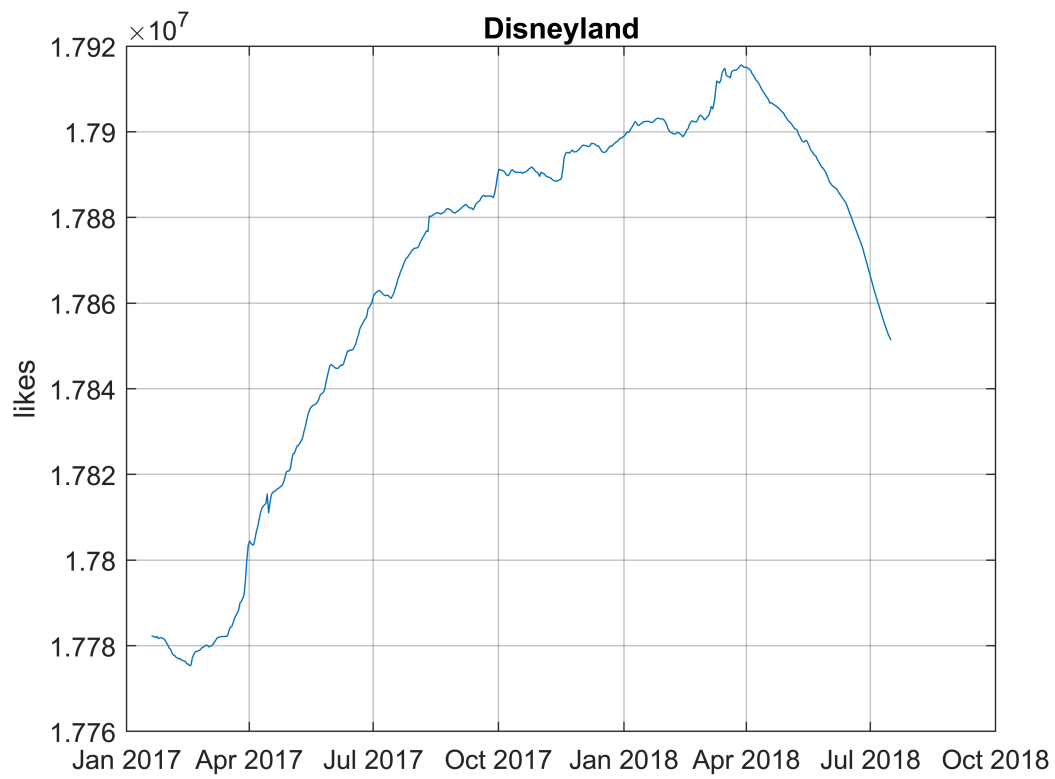
```
cindex = fb.username == 'valleyfair';
plot(fb.time(cindex),fb.likes(cindex))
title('Valleyfair')
ylabel('likes')
```

Valleyfair

(checkins vs. time, Jul 2015 – Jan 2018)

Disneyland and Valleyfair talking-about data over time

```
cindex = fb.username == 'Disneyland';
plot(fb.time(cindex),fb.talking_about_count(cindex))
title('Disneyland')
ylabel('talking-about-count')
```
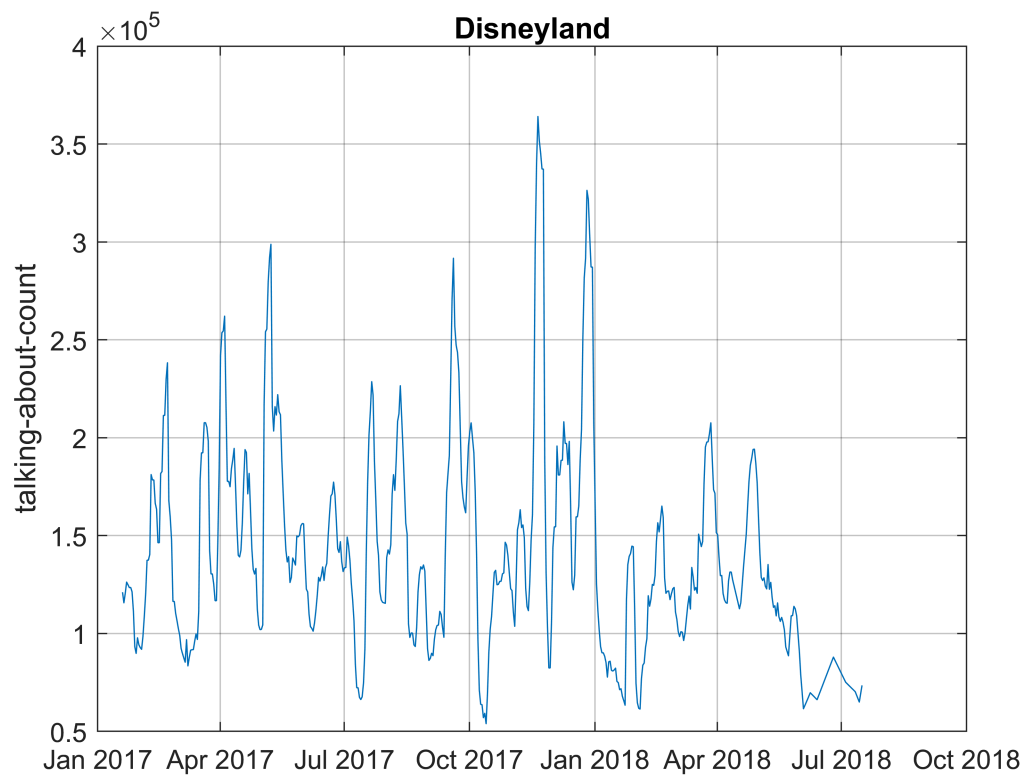
```
cindex = fb.username == 'valleyfair';
plot(fb.time(cindex),fb.talking_about_count(cindex))
title('Valleyfair')
ylabel('talking-about-count')
```

Valleyfair