



MOBILE APPS

“Quis custodiet ipsos custodes?”

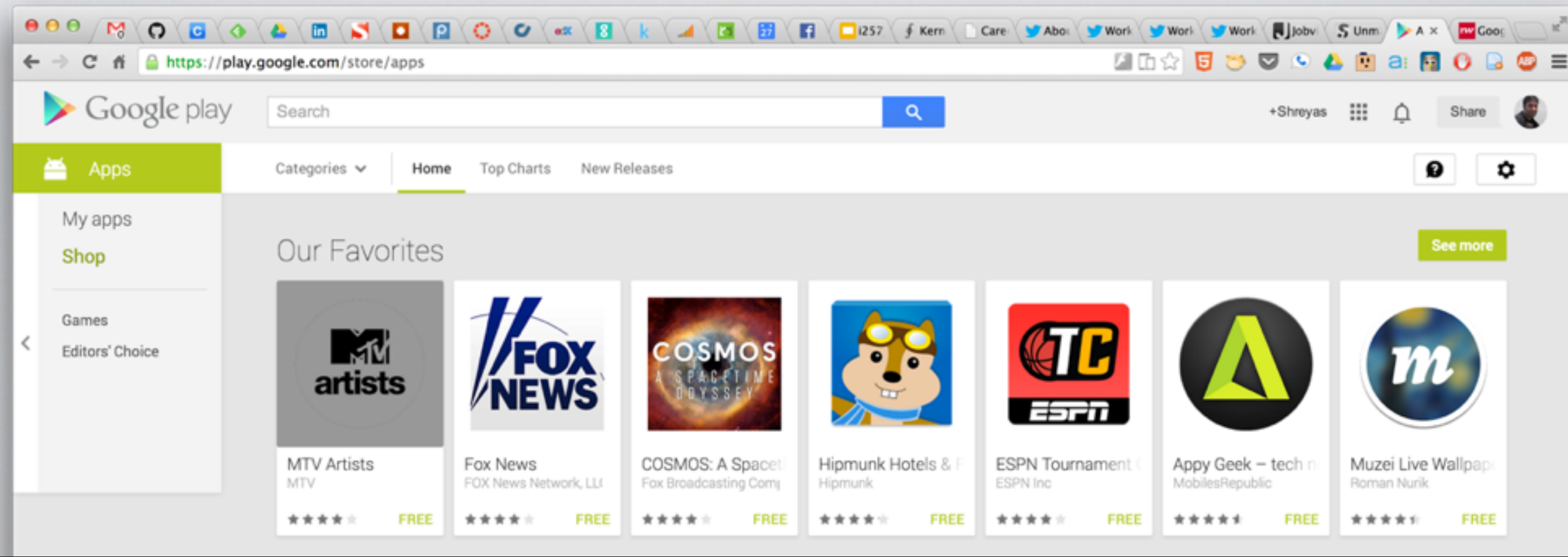
“Who guards the guardians”



Google Play Store

- ~**875,000** apps available
- ~**20,000** new apps added monthly
- Over **15 billion** downloads

source: appbrain



FTC / California AG

Currently looking for ways to find ***not just malware***, but also:

- **Deception:** "representation, omission, or practice that is likely to mislead"
- **Unfairness:** "injury to consumers with no countervailing benefit and that consumers could not avoid"

Not “Just” Malware

CONTACT

Brightest Flashlight app on Android caught selling user data to advertising companies

Posted on Dec 6 2013 - 8:00am by [Valerie Richardson](#)

« PREVIOUS | NEXT »



Android has had a terrible history with security and malware thanks numerous apps making its way to the **Google Play Store** with malicious intentions, despite Google's constant efforts to curb the nagging issues. And with user security and privacy being at the forefront of every person's life today, the FTC has just uncovered an app developer selling Android user data to third parties. This is a serious allegation and one which could cost dearly to the developers, but it is being said that the FTC has come to an agreement with the developers, although the details weren't disclosed to the public. So does this mean the





OBIDROID

“These aren’t the apps that you are looking for”

Aim to *sift through the android app store* applications and *build a flagging system* based on *key features* to indicate ***fair/unfair*** apps



Obidroid architecture



***compile a list** of apps from the app page*

scrape** the app **features

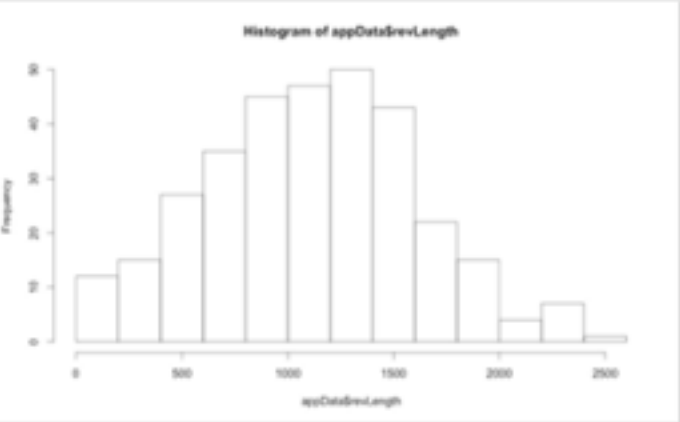
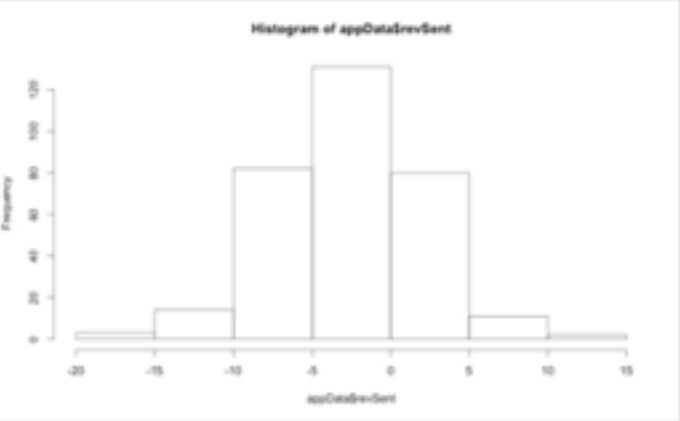
***push/pull the features** to/from a database*

*to **enable us to label** the apps based on features*

extract** features, **sentiment analysis** on sentences of review, **classify



Obidroid Features

Features Extracted	Feature Description	Feature Intuition	Feature Histogram
price	price of an app	Free apps might be more malware ridden	-
revLength	total sentences in all user reviews	longer the review, more coherent the user feeling about the app	
avgRating	average rating of the app	higher rated apps might be more reliable	-
hasPrivacy	whether the app has a privacy policy or not	FTC inspired	-
revSent	aggregate review sentiment	NLP inspired	
hasDeveloperEmail	app has an associated developer email	FTC inspired	-
hasDeveloperWebsite	app has an associated developer website	FTC inspired	-
countMultipleApps	app has multiple apps associated with it	self	-
installs	average install of each app	self	-



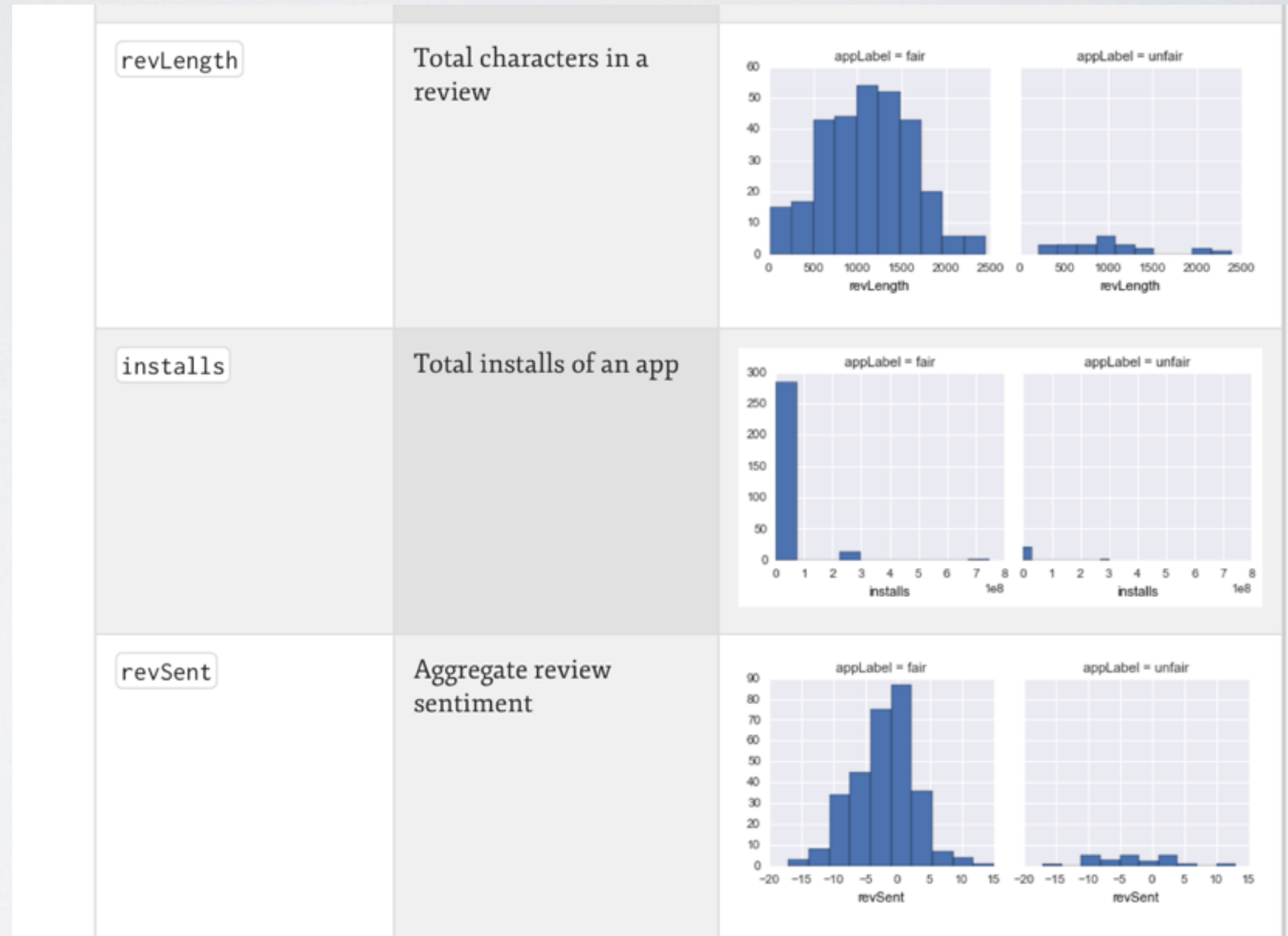
Obidroid Features

exclamationCount	count of exclamation for extreme reviews	NLP inspired	-
countCapital	count capitalized words in a review	NLP inspired	-
adjectiveCount	count the number of adjectives in	NLP inspired	-
positiveWordCount	count the number of positive words from a curated list	NLP inspired	-
negativeWordCount	count the number of negative words from a curated list	NLP inspired	-
unigrams like has(word)	presence of curated malindicator words	NLP inspired	-
bigrams	top 20 bigrams via likelihood ration measure	NLP inspired	-
trigrams	top trigrams based on raw frequency	NLP inspired	-



Obidroid Learning - *Unsupervised*

univariate analysis

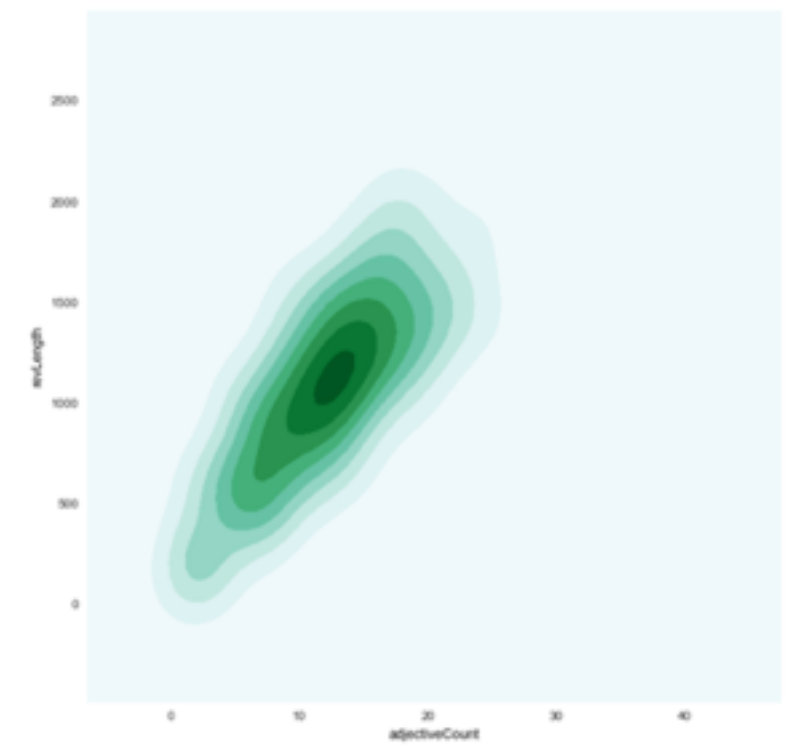
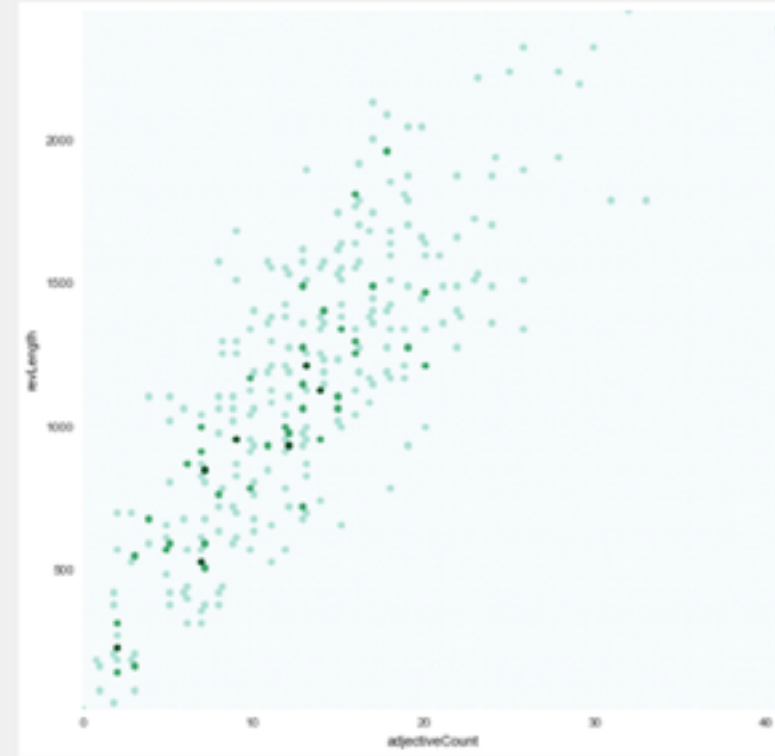




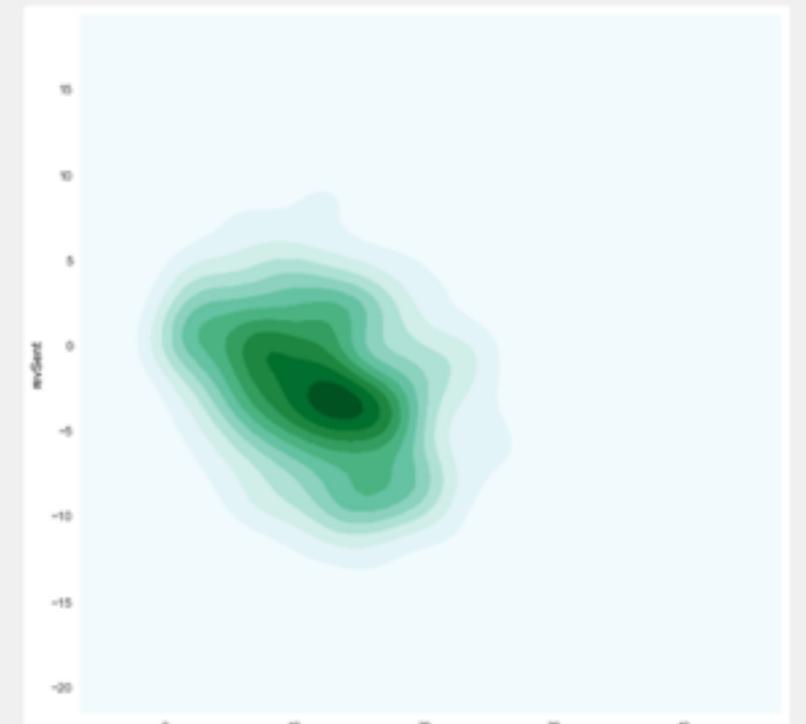
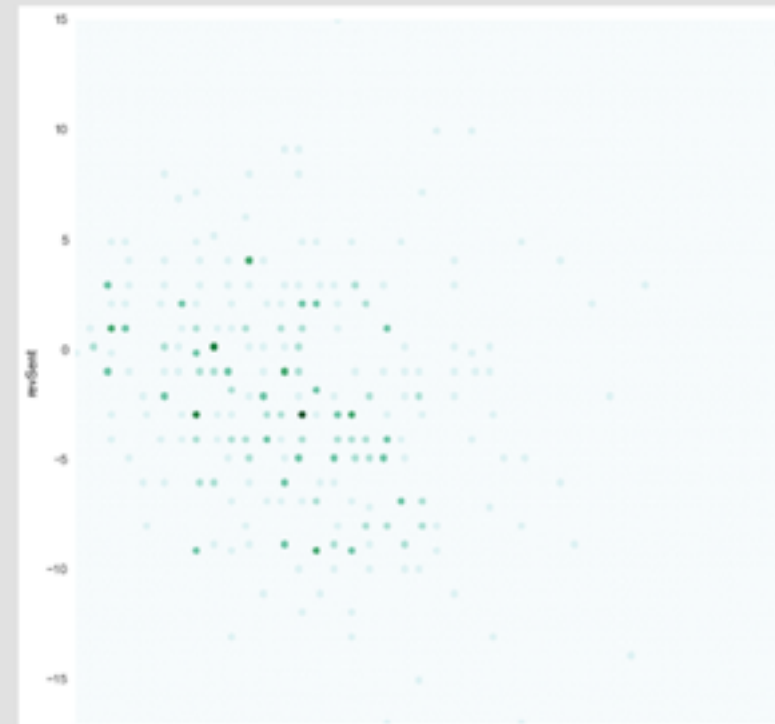
Obidroid Learning - *Unsupervised*

bivariate analysis

adjectiveCount x
revLength



adjectiveCount x
revSent

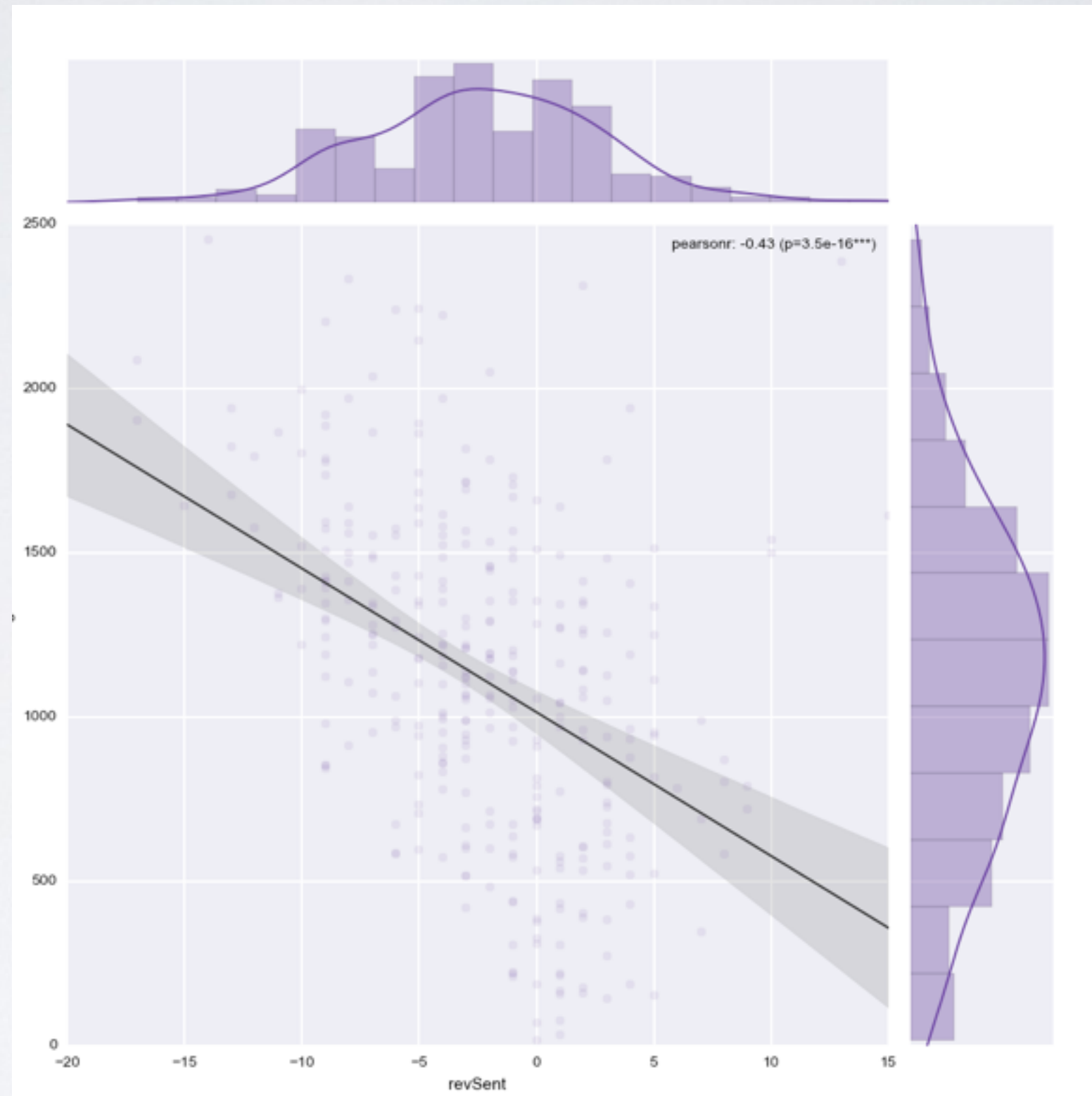




Obidroid Learning - *Unsupervised*

bivariate analysis

correlation

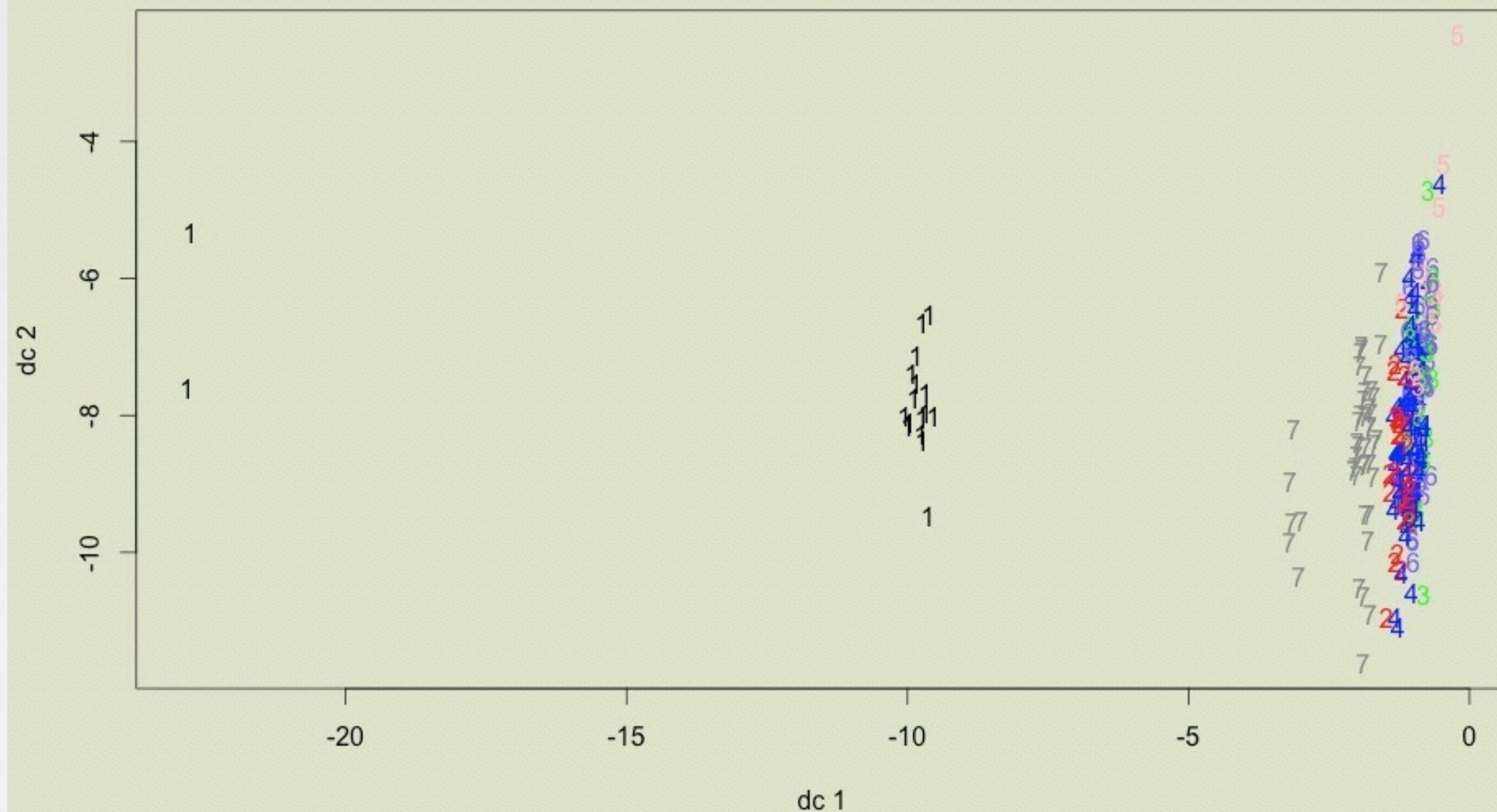




Obidroid Learning - *Unsupervised*

cluster analysis

k-means

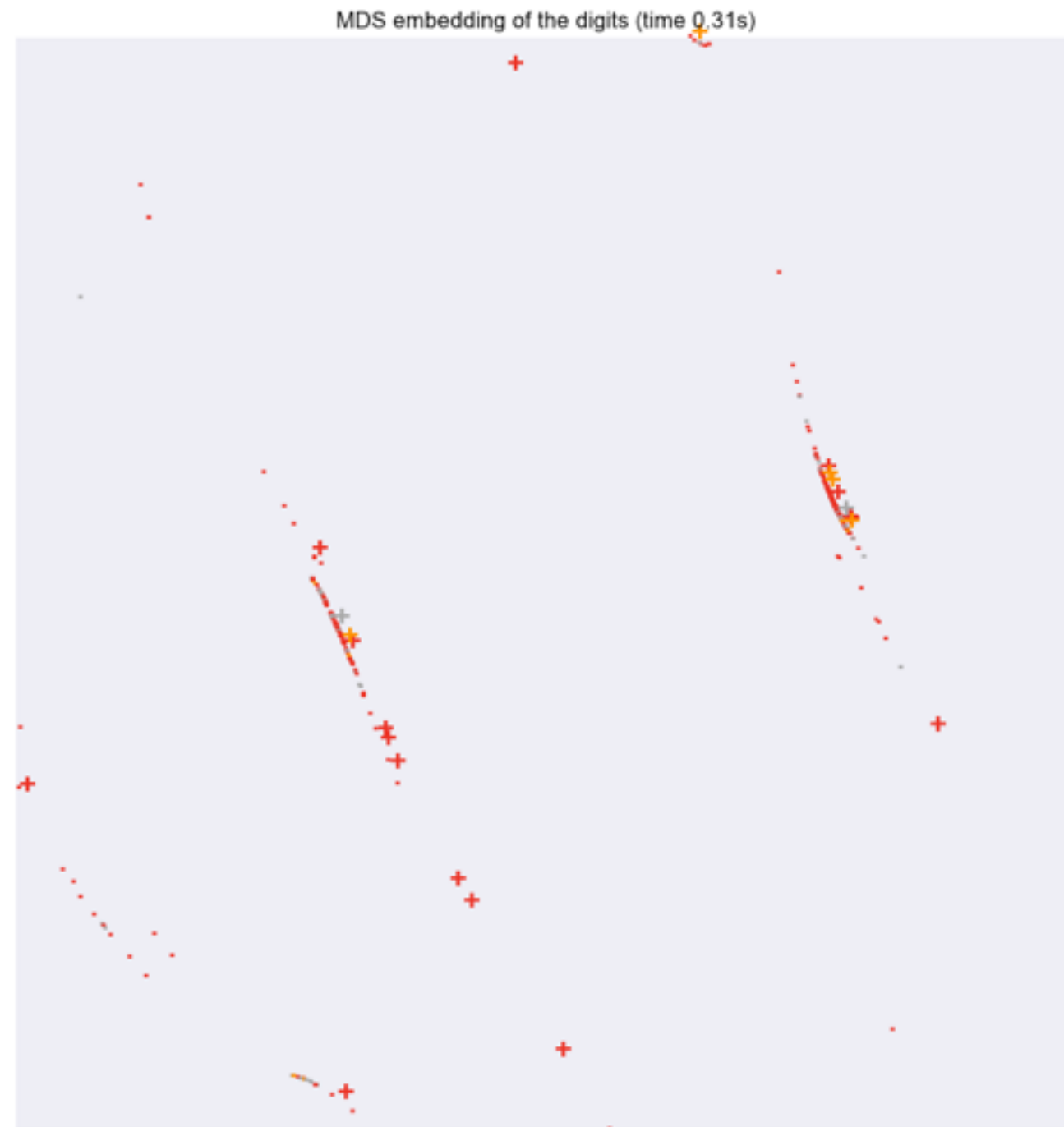




Obidroid Learning - *Unsupervised*

cluster analysis

mds





Obidroid Learning - *Supervised*

classifier

Naive Bayes

- Average Prediction Accuracy : 86.71875%
- Predictions in each fold: [0.8125, 0.90625, 0.84375, 0.90625]
- Overall Most Informative Features:
 - installs
 - for installs = 3000.0 the unfair : fair ratio was ~ 9 : 1
 - for installs = 30000.0 the unfair : fair ratio was ~ 6 : 1
 - for installs = 300000.0 the fair : unfair ratio was ~ 2 : 1
 - revSent
 - for revSent = -17 the unfair : fair ratio was ~ 8 : 1
 - for revSent = -10 the unfair : fair ratio was ~ 2 : 1
 - countCapital:
 - for countCapital = 9 the unfair : fair ratio was ~ 3 : 1
 - revLength:
 - for revlength = 800+ the unfair: fair ratio was ~ 2 : 1
 - avgRating:
 - ambiguous



Obidroid Learning - *Supervised*

more classifiers

```
# Supported classifier models
n_neighbors = 3
models = {
    'nb' : naive_bayes.GaussianNB(),
    'svm-l' : svm.SVC(),
    'svm-nl' : svm.NuSVC(),
    'tree' : tree.DecisionTreeClassifier(),
    'forest' : AdaBoostClassifier(tree.DecisionTreeClassifier(max_depth=1), algorithm="SAMME"),
    'knn-uniform' : neighbors.KNeighborsClassifier(n_neighbors, weights='uniform'),
    'knn-distance' : neighbors.KNeighborsClassifier(n_neighbors, weights='distance')
}
```