

DSI PROJECT 3: WEB APIs & CLASSIFICATION

DONE BY ZAINI CHIA

PRESENTATION OUTLINE

- **Problem Statement**
- **Data Cleaning**
- **Exploratory Data Analysis**
- **Modelling**
- **Interpretations**
- **Conclusions**

1. PROBLEM STATEMENT

In 2019, the global smartphone revenues amounted to a total of 400 billion US dollars. This is one of the world's biggest markets, with many companies heavily invested in it.

There are effectively only two smartphone platforms: iOS and Android. They are very different platforms though, and any insight as to how they work and what customers like about each platform, can be very valuable to hardware companies, software and services companies, and many more.

In this project, I will attempt to understand the similarities and differences between these two dominant smartphone platforms.



2. DATA CLEANING

	author_fullname	title	selftext	subreddit
0	t2_1unjxir5	Samsung trade in program is back again for Bal...		Android
1	t2_56a46frx	Vivo unveils Origin Os - Will it be less polar...		Android
2	t2_b8kyv	Nokia 2.4 hands-on - GSMArena		Android
3	t2_cqcpv2h	Are high-speed/high-capacity (A1, A2, U3) micr...	After recently acquiring a 1TB microSD card in...	Android
4	t2_5l3x1fsc	Android App Bundles could have a drawback for ...		Android

- A lot of posts with a title and just a link/image/video in the main body
- Combined title + selftext

2. DATA CLEANING

- Date goes from **26/11** to **22/10**, and then to **26/11** again
- Duplicate posts

```
iPhone_posts['created_utc']
```

0	1.606363e+09
1	1.606353e+09
2	1.606351e+09
3	1.606348e+09
4	1.606348e+09

```
time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(iPhone_posts['created_utc'][0]))
```

'2020-11-26 12:04:30'

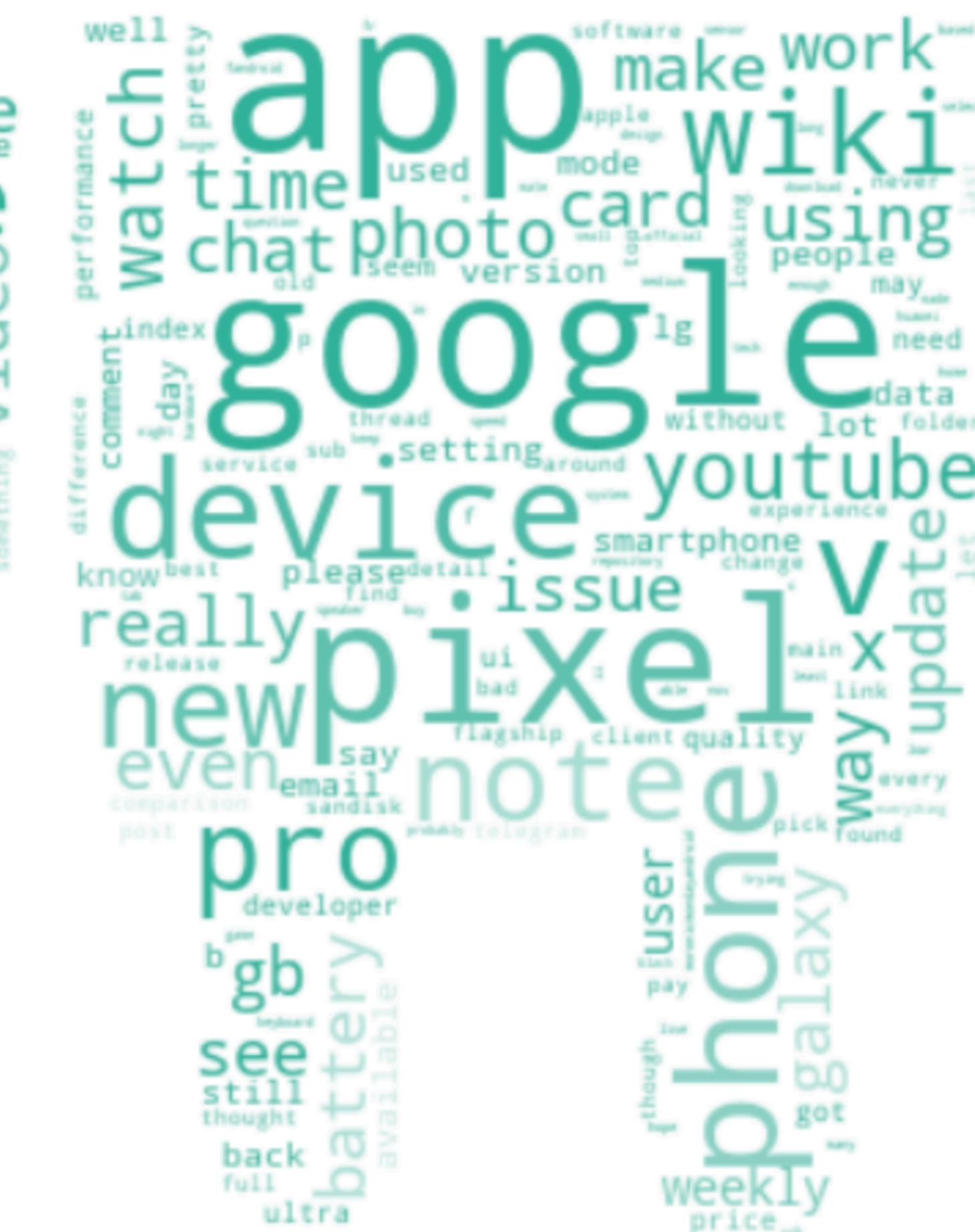
```
time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(iPhone_posts['created_utc'][940]))
```

'2020-10-22 02:24:37'

```
time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(iPhone_posts['created_utc'][950]))
```

'2020-11-26 04:07:22'

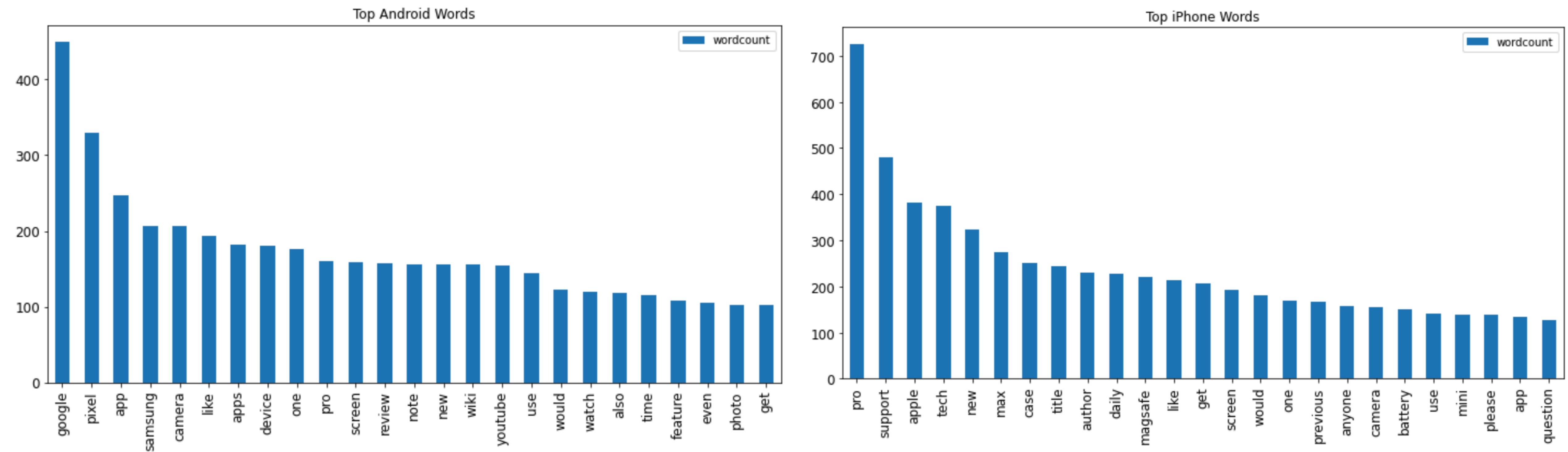
3. EXPLORATORY ANALYSIS



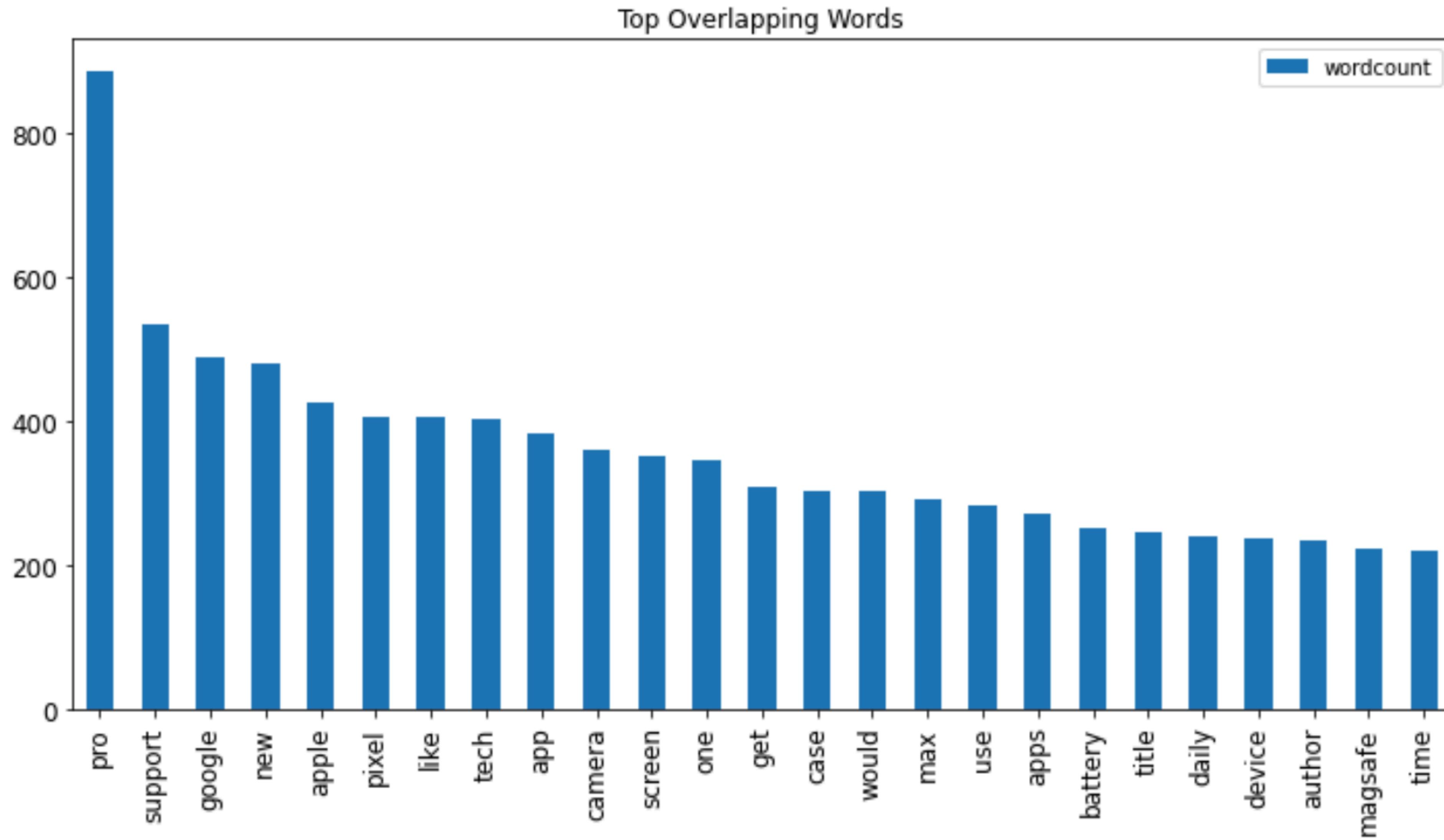
3. EXPLORATORY ANALYSIS



ANDROID VS IPHONE



OVERLAPPING WORDS



4. MODELING

```
pipe_tv_lr = Pipeline([
    ('tvec', TfidfVectorizer(stop_words = updated_stops)),
    ('lr', LogisticRegression())
])
```

```
pipe_params = {
    'tvec__max_features': [2000, 3000, 4000],
    'tvec__min_df': [2, 3, 4],
    'tvec__max_df': [0.85, 0.88, 0.92],
    'tvec__ngram_range': [(1, 2), (1, 3)],
    'lr_C': [0.5, 0.8, 1.0]
}
```

```
gs_tv_lr = GridSearchCV(pipe_tv_lr, param_grid = pipe_params,
                        cv = 5, verbose = 1)
```

```
gs_tv_lr.fit(X_train, y_train)
```

Fitting 5 folds for each of 162 candidates, totalling 810 fits

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 810 out of 810 | elapsed: 3.8min finished

```
tv_lr_model = gs_tv_lr.best_estimator_
tv_lr_model.score(X_train, y_train)
```

0.9640287769784173

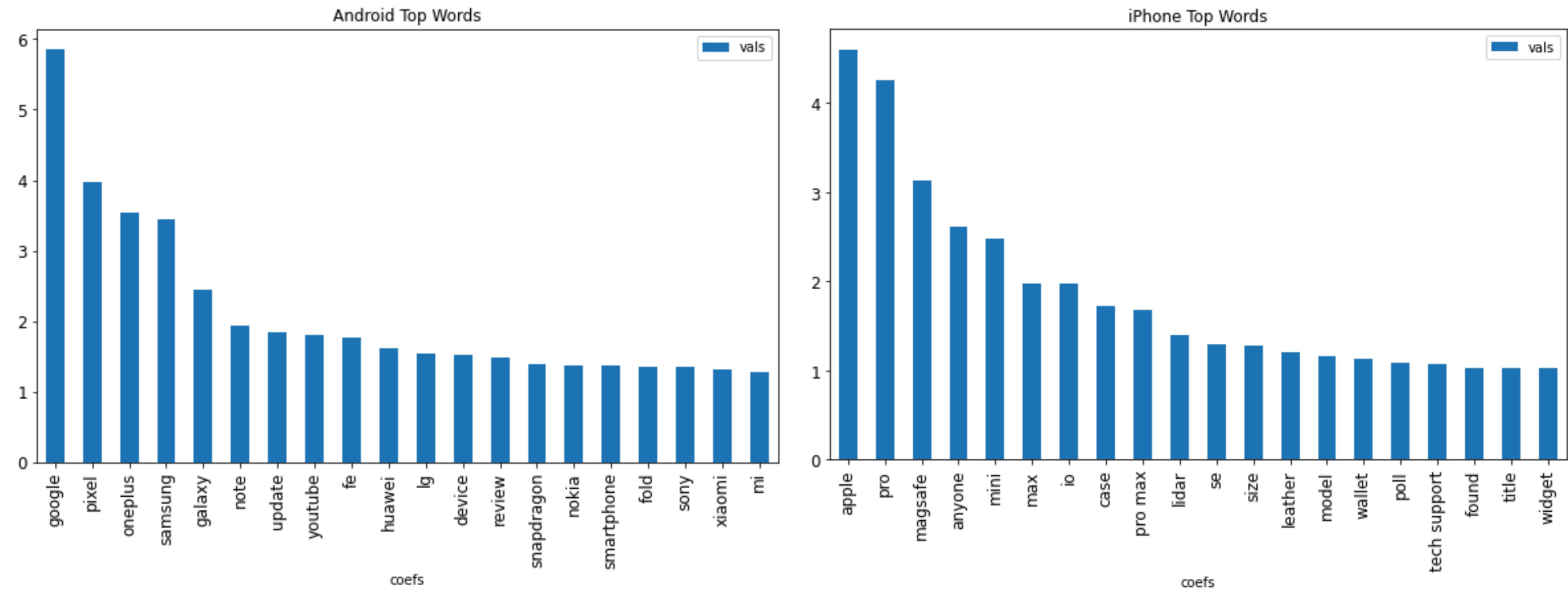
```
tv_lr_model.score(X_val, y_val)
```

0.894484412470024

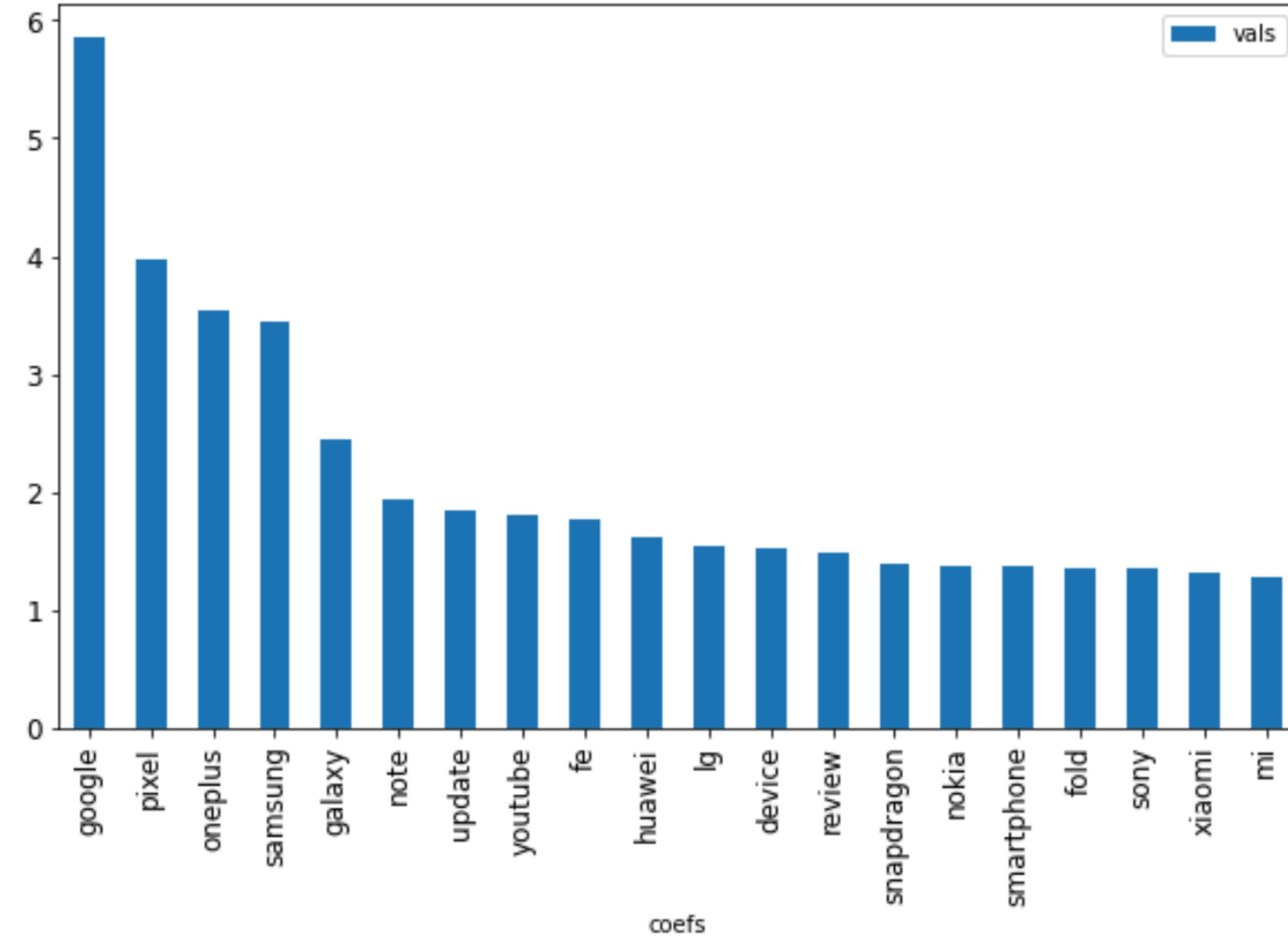
```
trained_model.score(X_test, y_test)
```

0.944

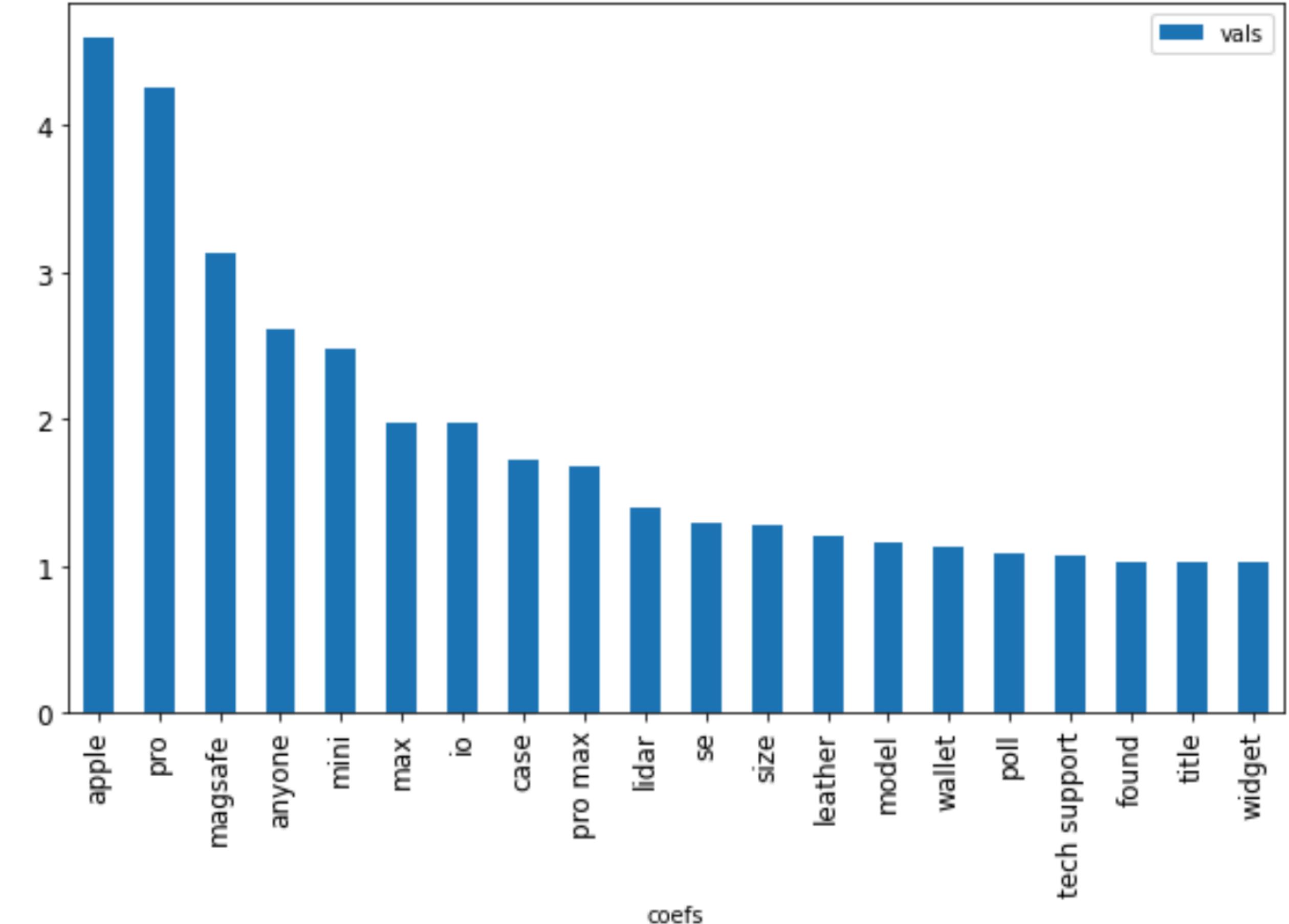
5. INTERPRETATIONS



Android Top Words

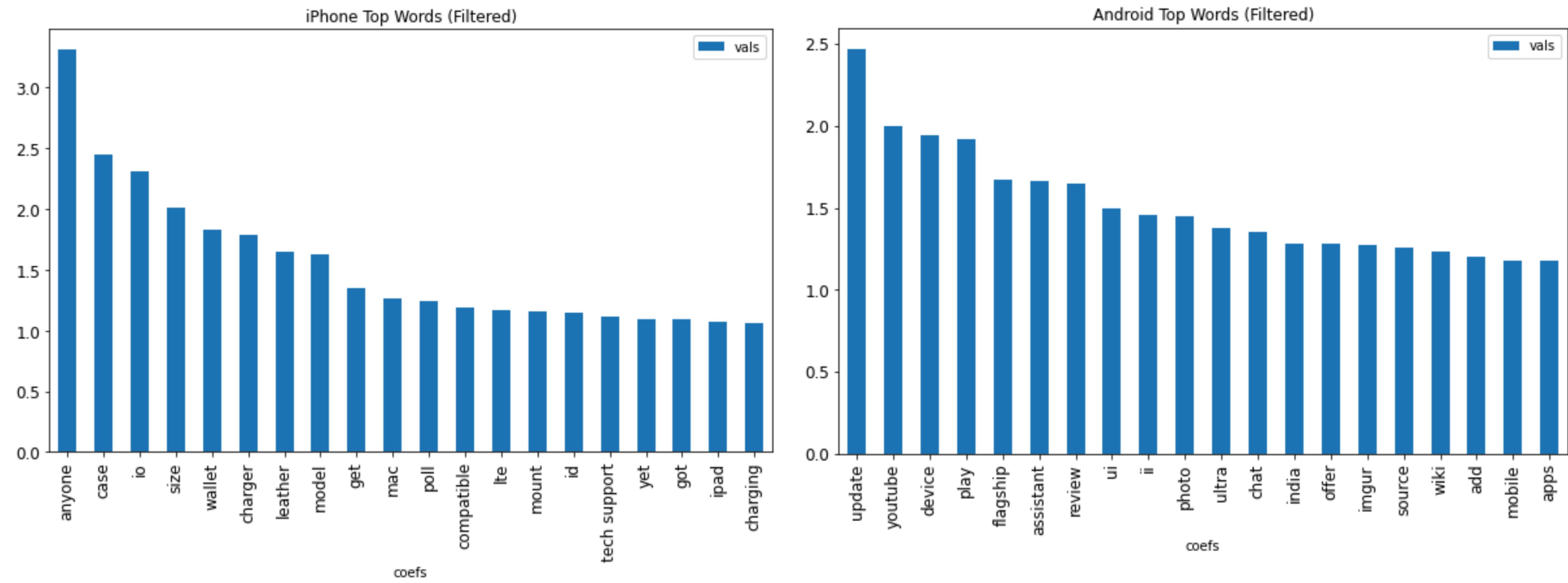


iPhone Top Words

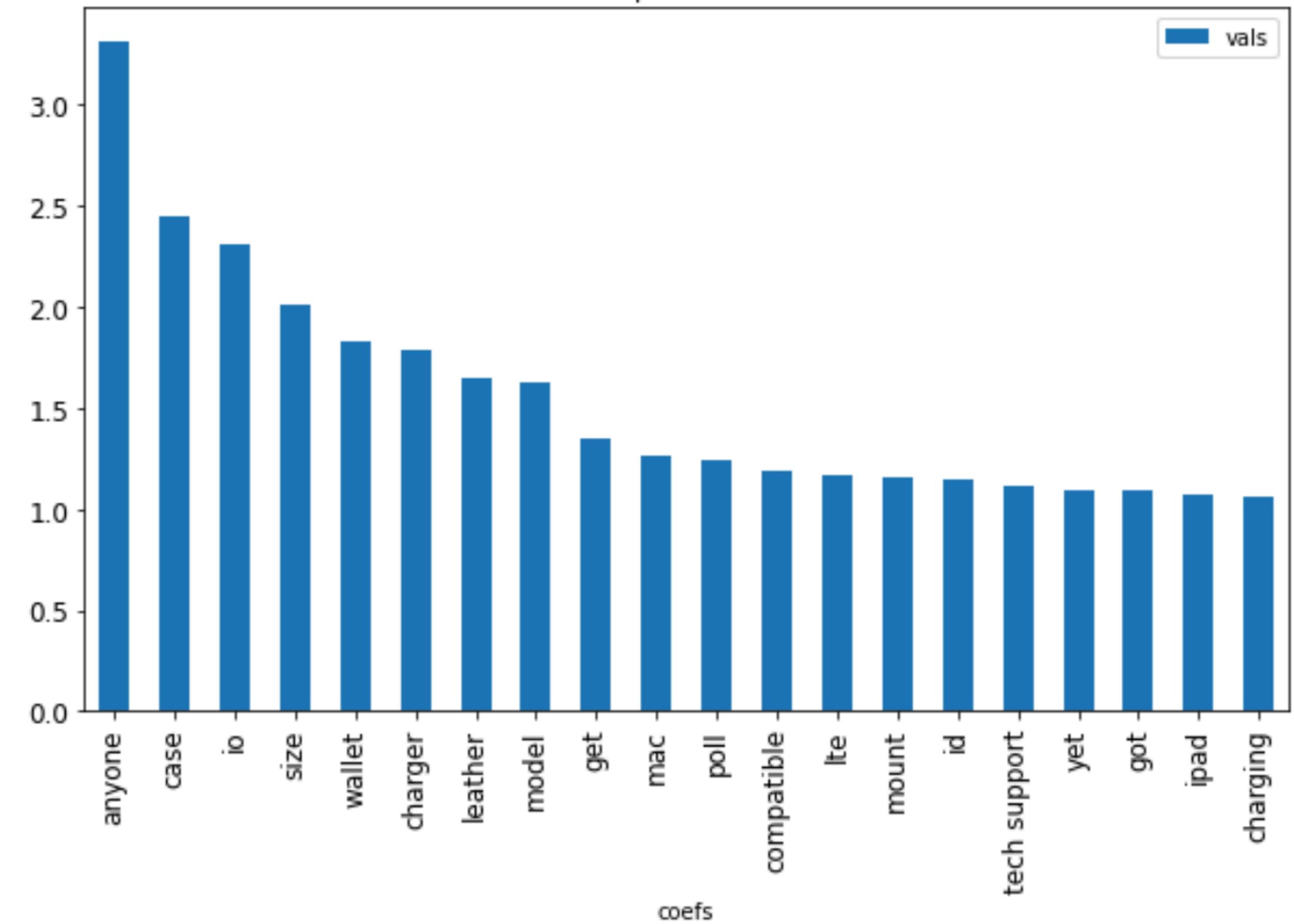


- A lot of brand names and specific feature names
- Google, Pixel, Apple, LiDAR, Magsafe, etc
- Remove to 'dig deeper' into the data

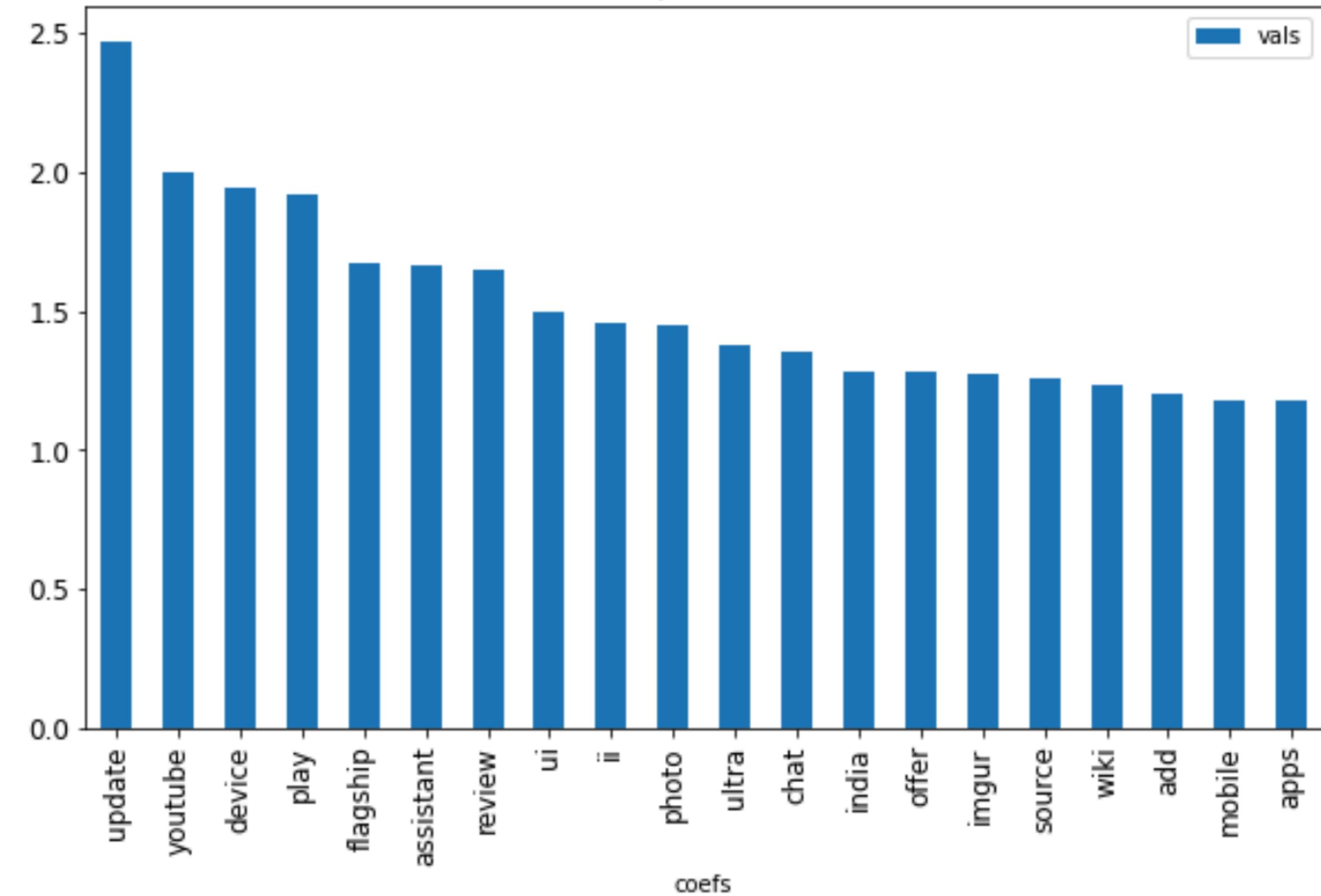
5. INTERPRETATIONS (FILTERED)



iPhone Top Words (Filtered)

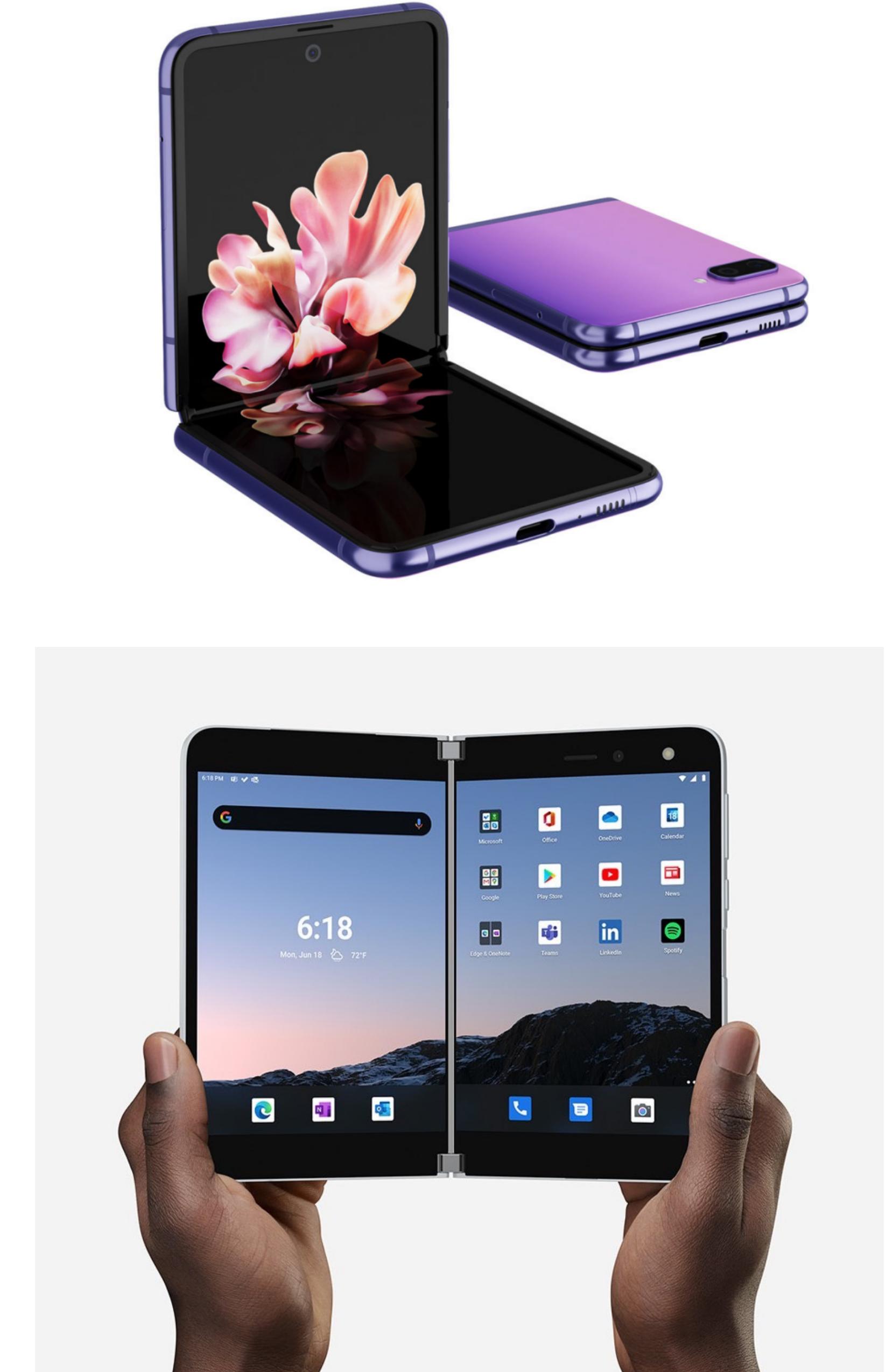
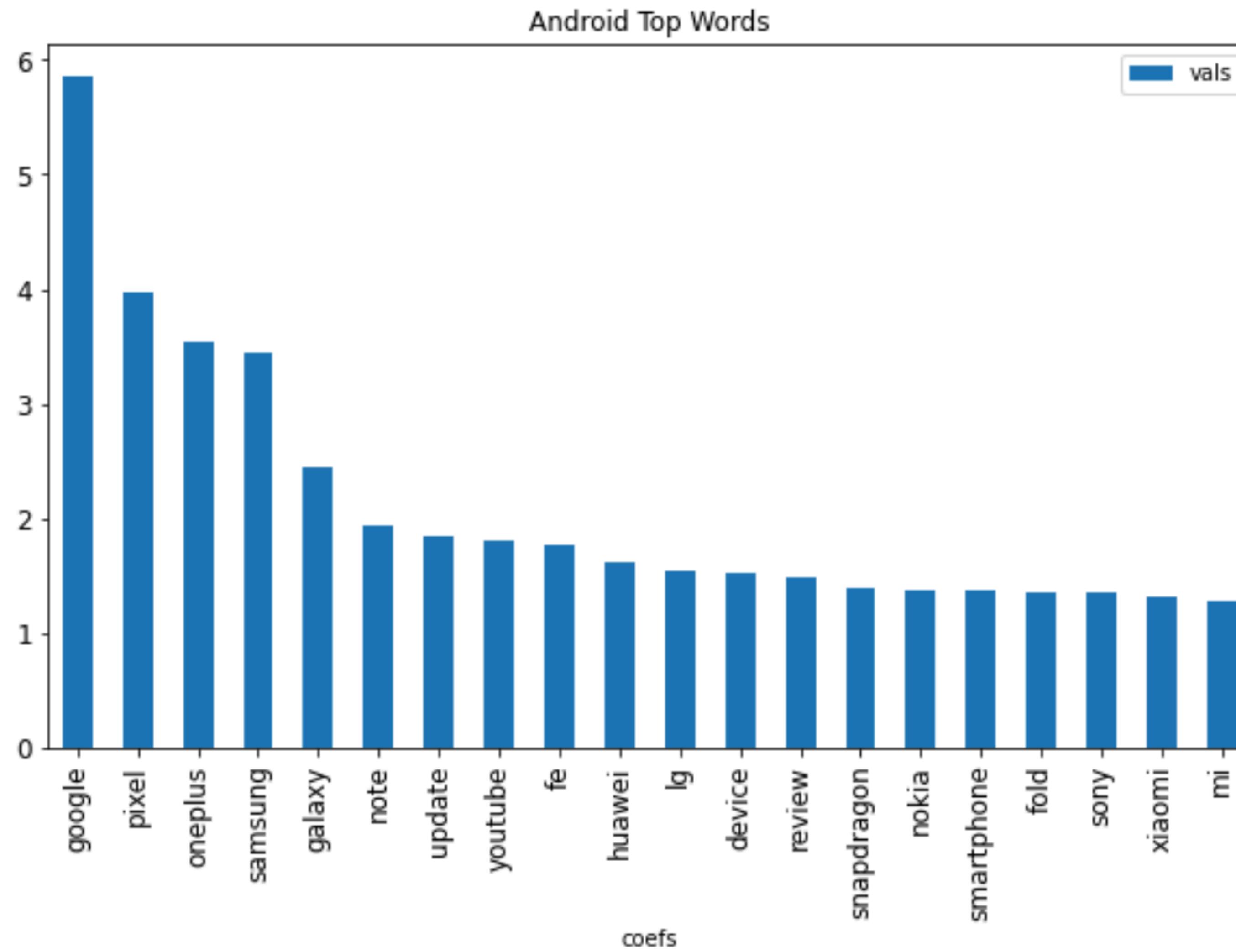


Android Top Words (Filtered)



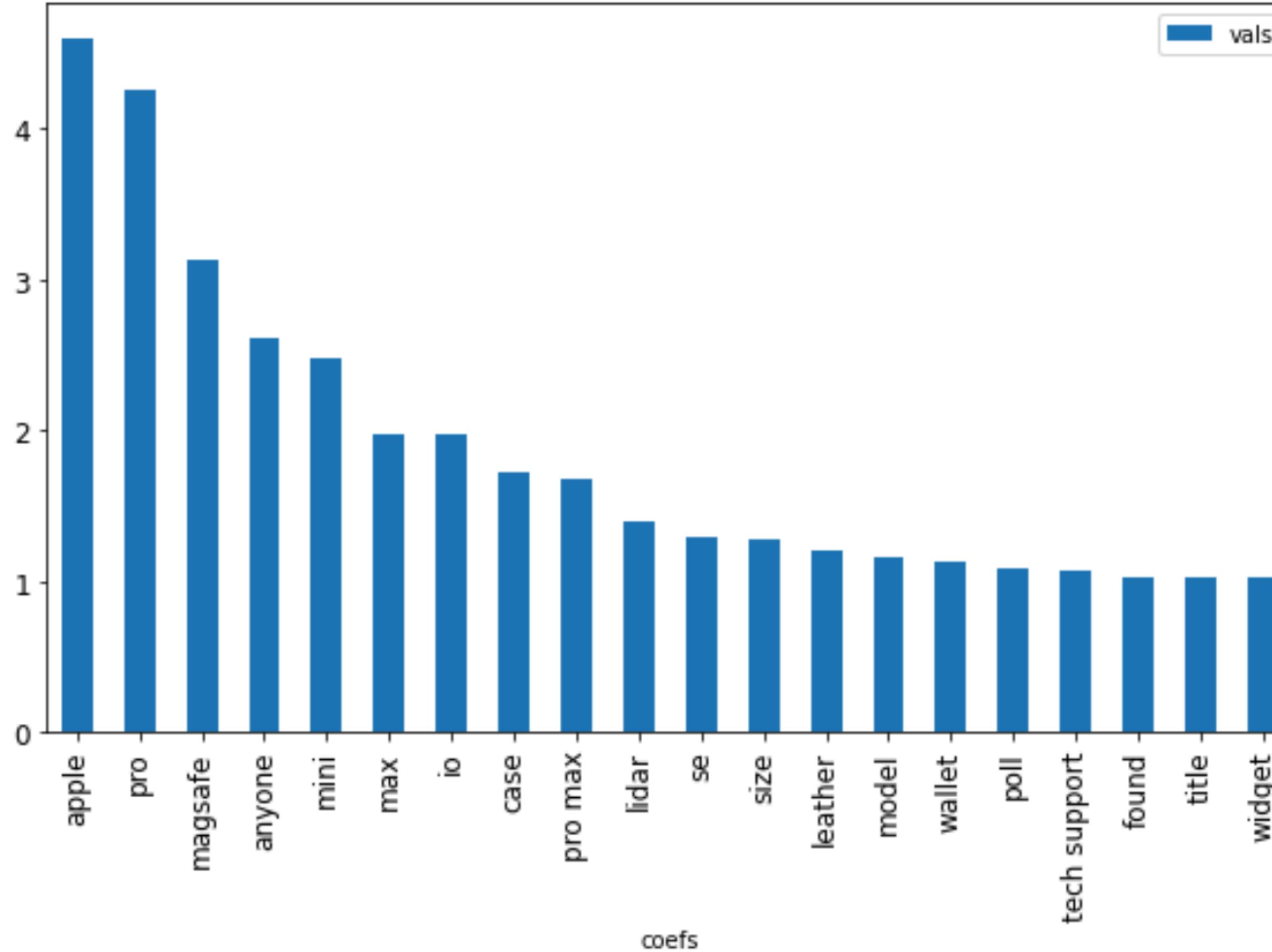
- iPhone fans talk about ecosystem: ‘Mac’, ‘iPad’, ‘iOS’, accessories, tech support
- Android fans are more technical: ‘update’, ‘assistant’, ‘UI’(User-Interface), ‘flagship’

6. CONCLUSIONS



6. CONCLUSIONS

iPhone Top Words



And boom.
You own a film studio.

Now you can record 4K HDR video with Dolby Vision. Edit in Photos or iMovie. Then zap it to your TV using AirPlay⁷ to see all that color and detail. It's a start-to-finish cinematic experience — found only on iPhone. Which comes in handy when your dog does something adorable.

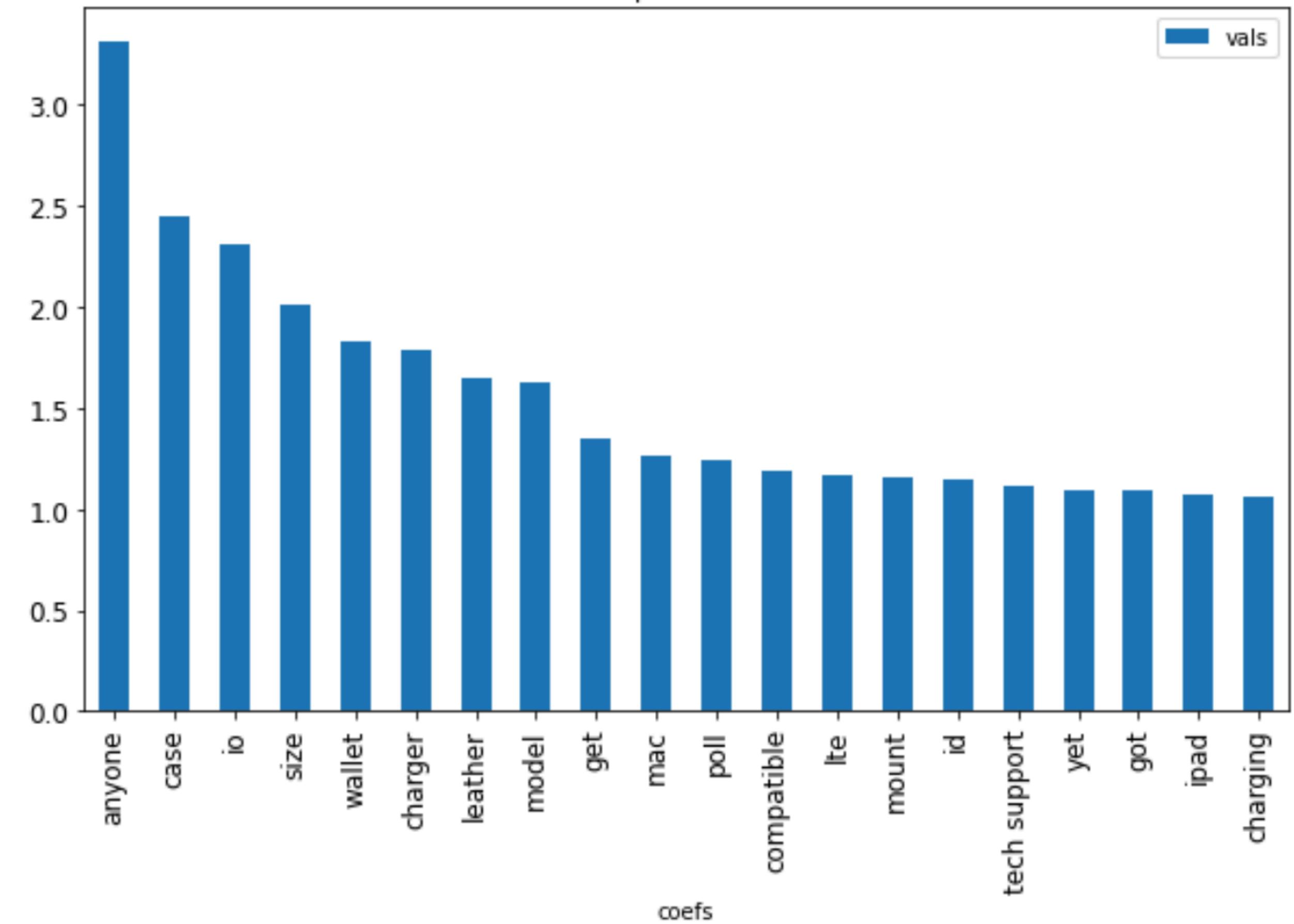
First camera ever to
record in Dolby Vision

[Find out how](#)

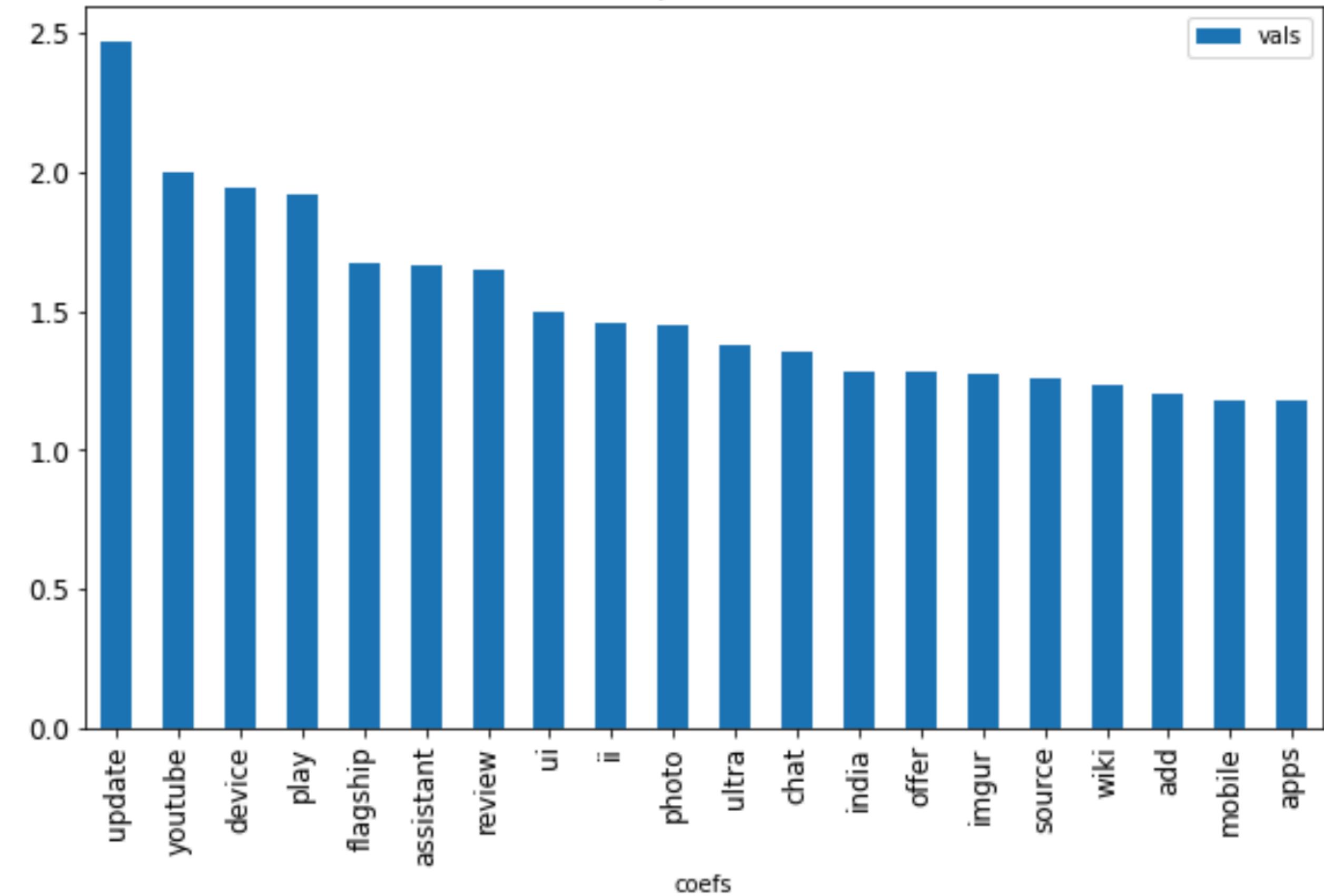
Kicks glass.

Ceramic Shield, tougher than
any smartphone glass

iPhone Top Words (Filtered)



Android Top Words (Filtered)



- iPhone fans talk about ecosystem: ‘Mac’, ‘iPad’, ‘iOS’, accessories, tech support
- Android fans are more technical: ‘update’, ‘assistant’, ‘UI’(User-Interface), ‘flagship’

THANK YOU!

