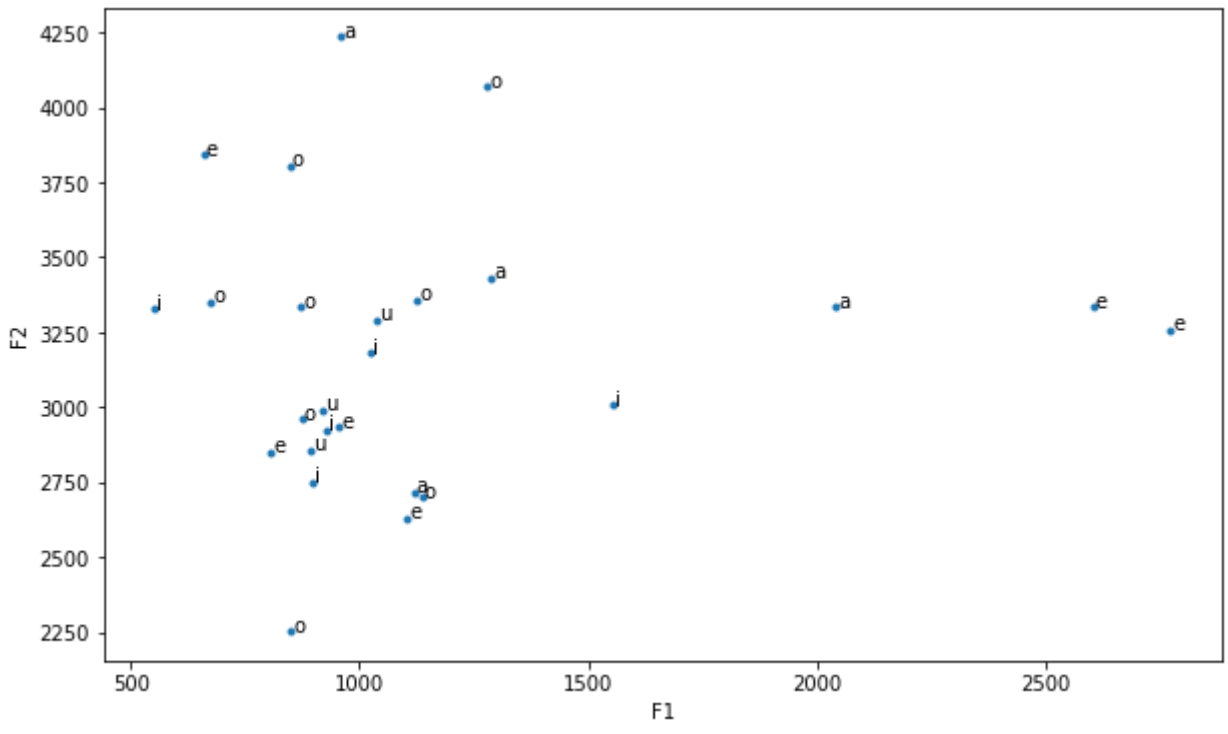
In [1]:
```
# data

# (label, f1, f2)
# the quick brown fox jumps merrily over the lazy dog
data = [('i', 1553, 3009), ('o', 874, 2963), ('o', 872, 3336), ('u', 917, 2992), ('o', 1280, 4067), ('a',
1289, 3433), ('i', 551, 3329), ('o', 1125, 3359), ('e', 807, 2850)]
# lake, tree, pine cone, cool brown soil
data+= [('a', 2042, 3335), ('e', 2603, 3334), ('i', 898, 2751), ('o', 675, 3352), ('o', 849, 2256)]
# fat pet pig got drunk, took walk
data+= [('a', 1119, 2717), ('e', 956, 2935), ('i', 1025, 3183), ('o', 1136, 2702), ('u', 893, 2857), ('o'
, 848, 3804), ('a', 959, 4234)]
# where my cow boy here say hello sure
data+= [('e', 1104, 2630), ('i', 926, 2925), ('e', 2772, 3259), ('e', 660, 3843), ('u', 1037, 3291)]
```

In [4]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.cluster import KMeans

dataF = [(x[1], x[2]) for x in data]
dataX, dataY = [x[0] for x in dataF], [x[1] for x in dataF]

# plot graph
fig, ax = plt.subplots(figsize=(10,6))
ax.scatter(dataX, dataY, marker='.')
plt.xlabel('F1'), plt.ylabel('F2')

for phoneme, x, y in data:
    ax.annotate(phoneme, (x+5, y))
```



In [3]:
```
# perform k-means
kmeans = KMeans(n_clusters=5, init='k-means++', max_iter=300, n_init=10, random_state=0)
kmeans.fit(dataF)
pred_y = kmeans.fit_predict(dataF)

# plot graph
fig, ax = plt.subplots(figsize=(10,6))
ax.scatter(dataX, dataY, marker='.')
plt.xlabel('F1'), plt.ylabel('F2')
for phoneme, x, y in data:
    ax.annotate(phoneme, (x+5, y))

ax.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=60, c='red')
```

Out[3]: <matplotlib.collections.PathCollection at 0x7fbb88740790>