INFO7374 Algorithmic Digital Marketing

Summary	In this assignment, we implemented all the three implementations. And create two respective apps to fulfill the two modes.
Author	Yifu Liu and Pengfei He

Data Preprocessing

Sampling Data:

We used the **bson_and_image preprocessing.ipynb** to sample dataset, to generate sample images, as well as to have an insight of the category-id

```
pool.join()

print('Images saved at %s' % images_dir)
print('Products: \t%d\nCategories: \t%d\nPictures: \t%d' % (pro

file = open(os.path.join(base_dir, 'retrained_labels.txt'), 'w'

rootdir_glob = images_dir + '/**/*'
folder_list = [f for f in iglob(rootdir_glob, recursive=True) i
for folder in folder_list:
    category = folder.split('/')[-1]
    file.write(category + '\n')

file.close()

print('"retrained_labels.txt" saved at %s' % base_dir)
```

```
82it [00:00, 101.84it/s]
Images saved at /Users/check4068/images
Products:       2
Categories:     2
Pictures:       2
"retrained_labels.txt" saved at /Users/check4068
```

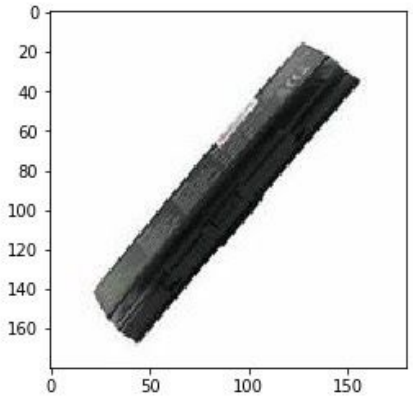
```
[5]: prod_to_category
```

[5]:

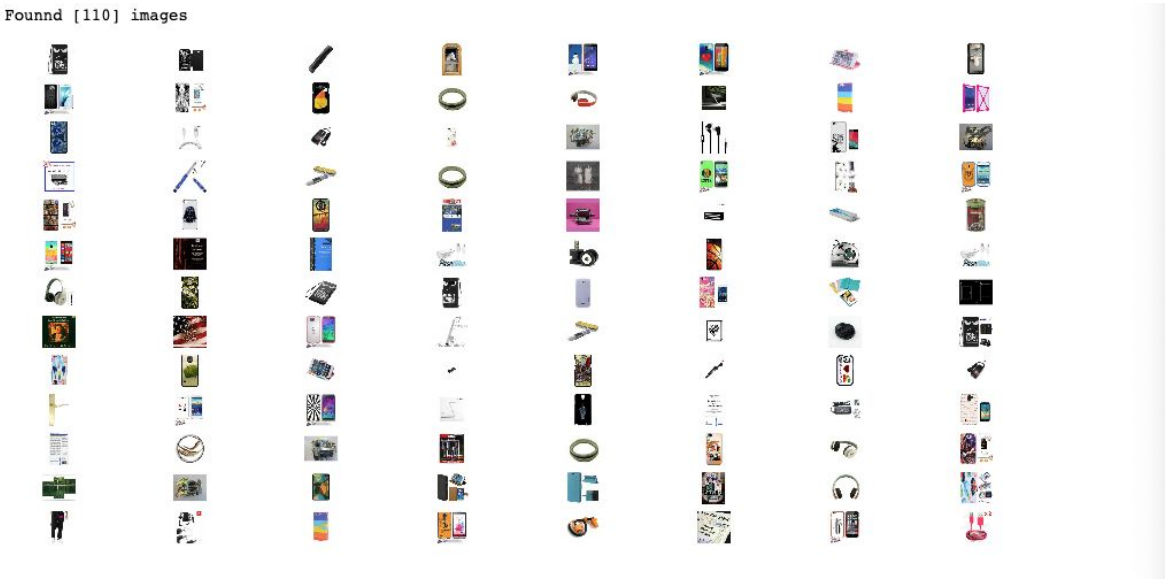
_id	category_id
0	1000010653
1	1000010653
2	1000004079
3	1000004141
4	1000015539
...	...
95	1000010653
97	1000010683
98	1000010667
99	1000014053
101	1000004085

82 rows × 1 columns

```
[6]: plt.imshow(picture);
```



Sample Images to Use:

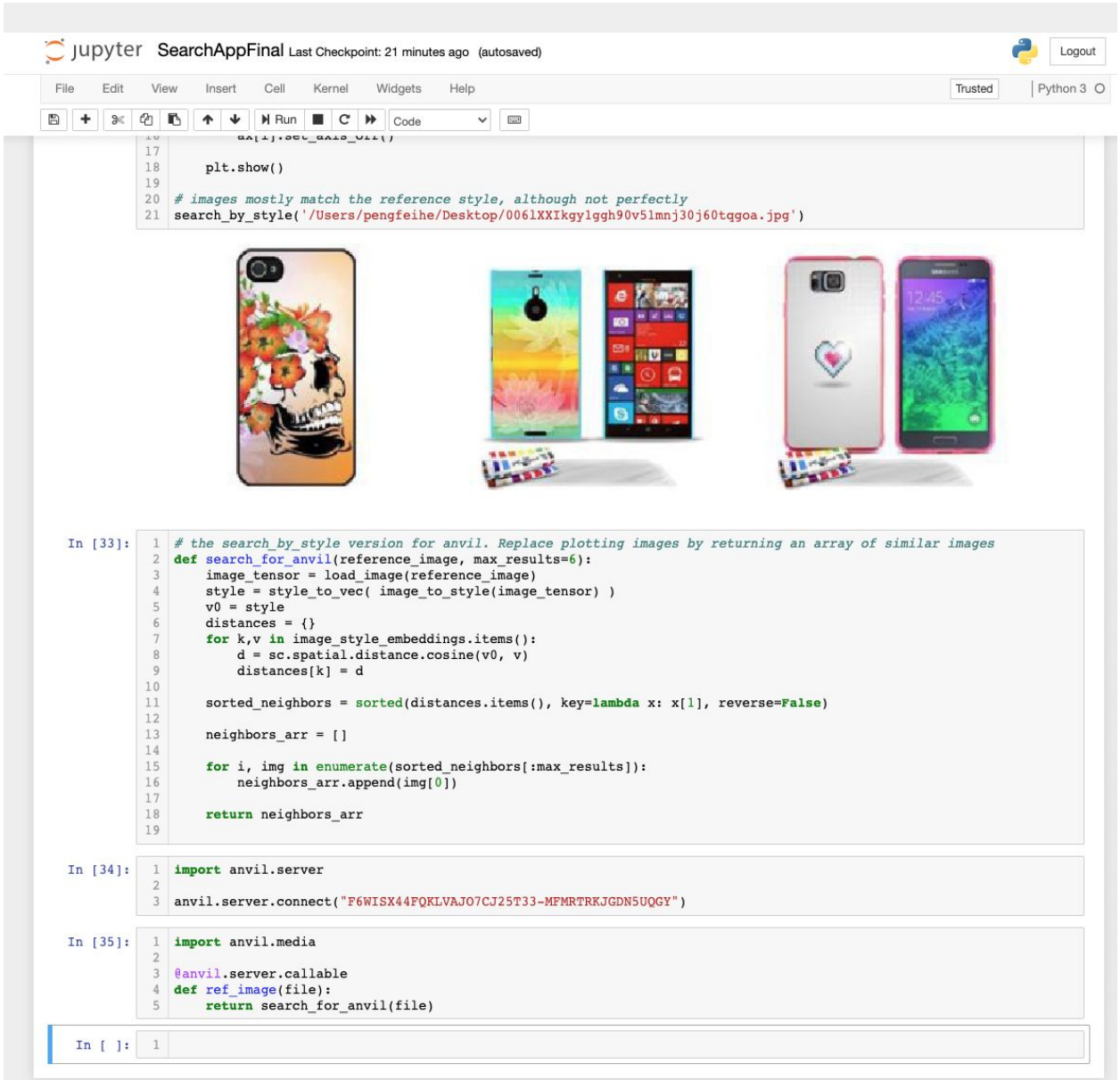


Implement of Version 1 (App mode 2)

We used the **first implement version** to create an app **fulfill the mode two**, which allows user to upload a new image and return k images similar to it. This app is published on **Anvil cloud**. The public url is **bitter-general-top.anvil.app**

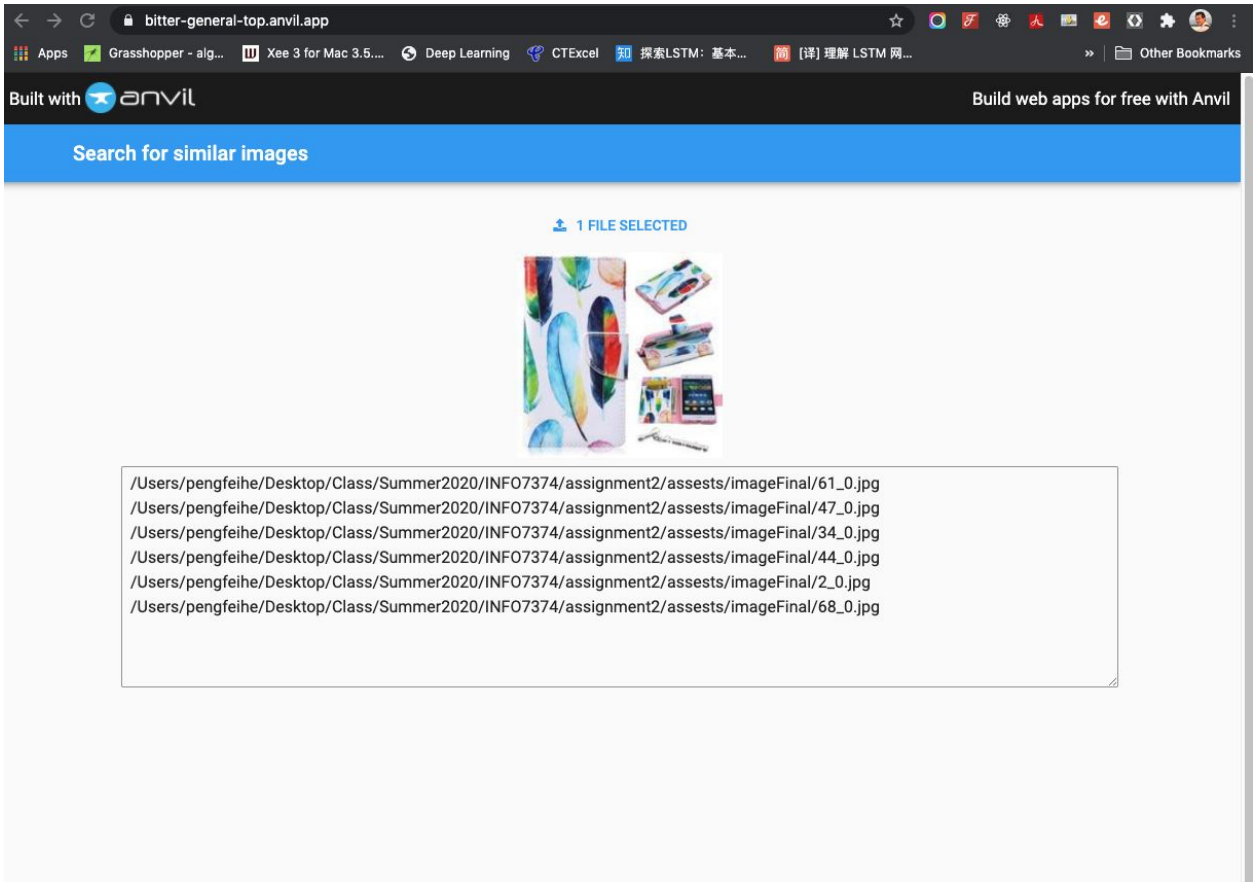
Implementing version_1 locally:

We modified the search_by_style function, by generating a new style for each new uploaded image, and then match this new style to the existing style array. In this way, we can allow users to **randomly upload any images**, and still can return the best similar images.



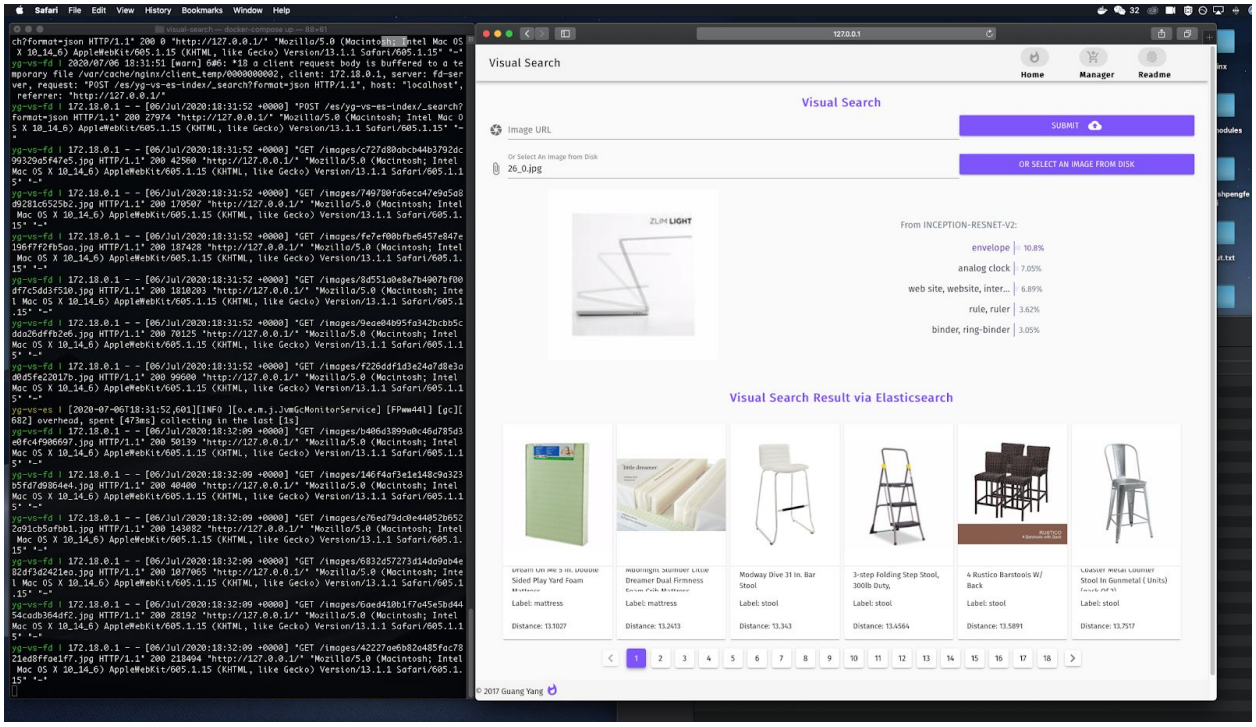
Deployed to Anvil Cloud (url:bitter-general-top.anvil.app)

We used Anvil to turn the jupyter notebook into a web app. And deployed it into Anvil Cloud server.



Implement of Version 2 (Run on docker)

Since the reference app for the **second implement version** is already a fully developed search app. We just **run it on docker** to use it.

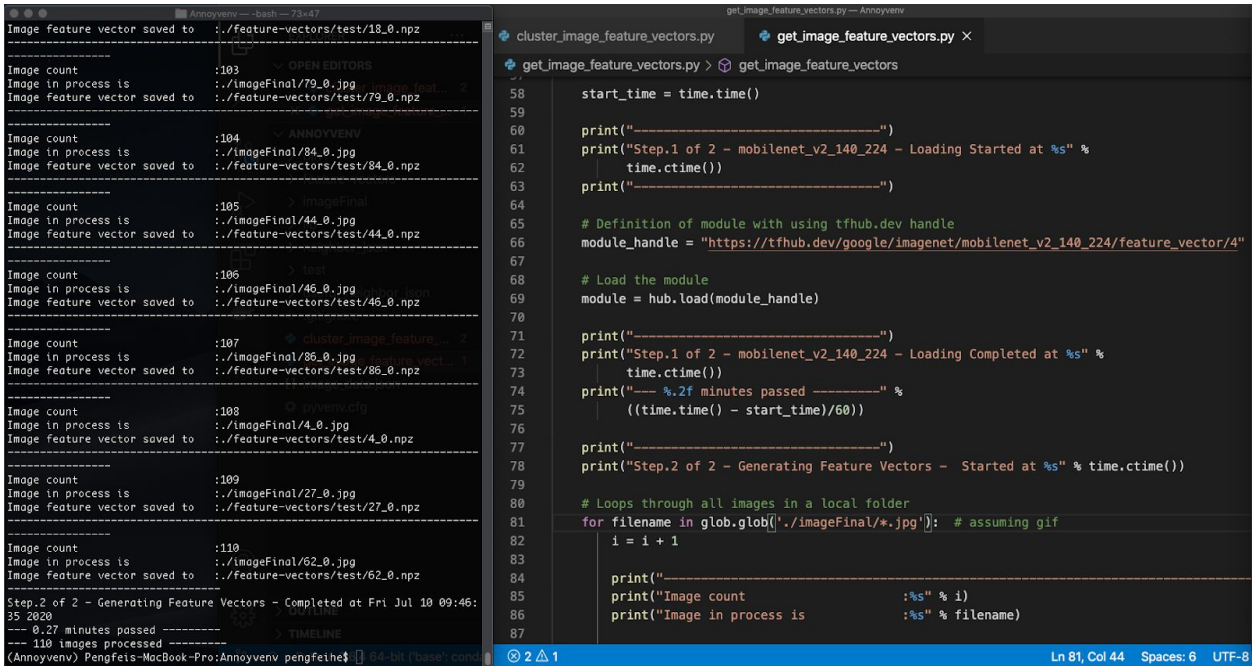


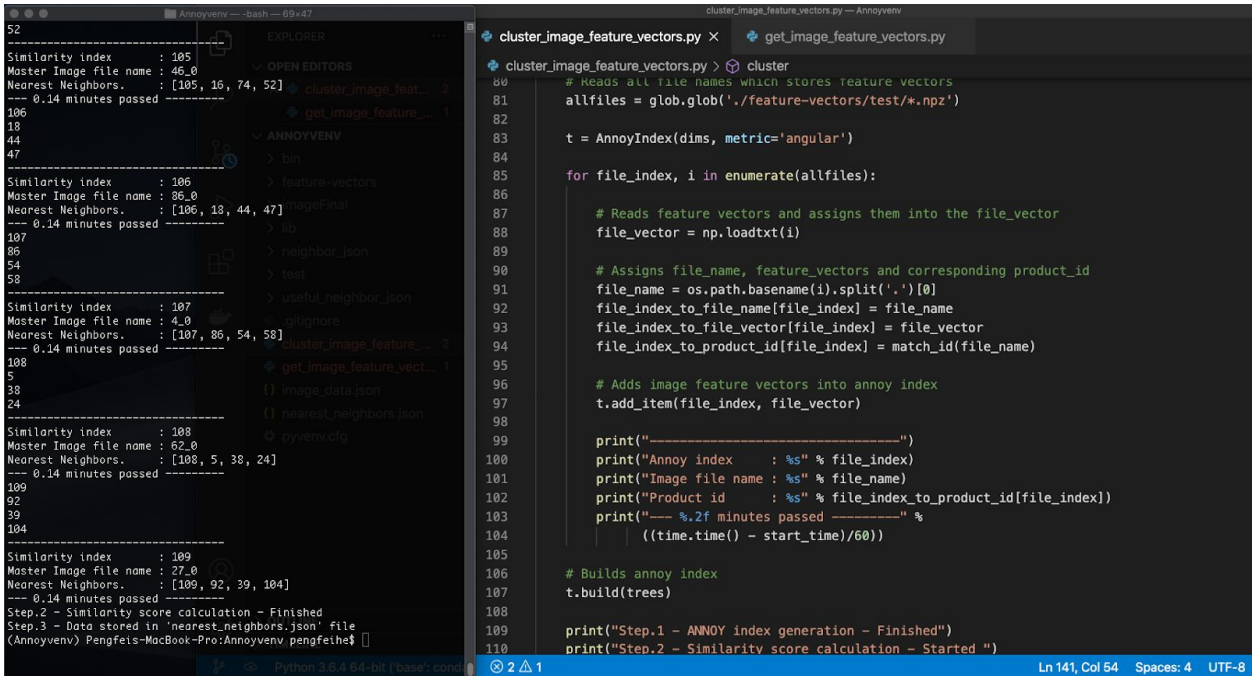
Implement of Version 3 (App mode 1)

We used the **third implement version** to create an app **fulfill the mode one**, which allows the user to select one image and get k similar images based on **elasticsearch**. Our web app is developed using django framework.

Generate N-Nearest Neighbor json file

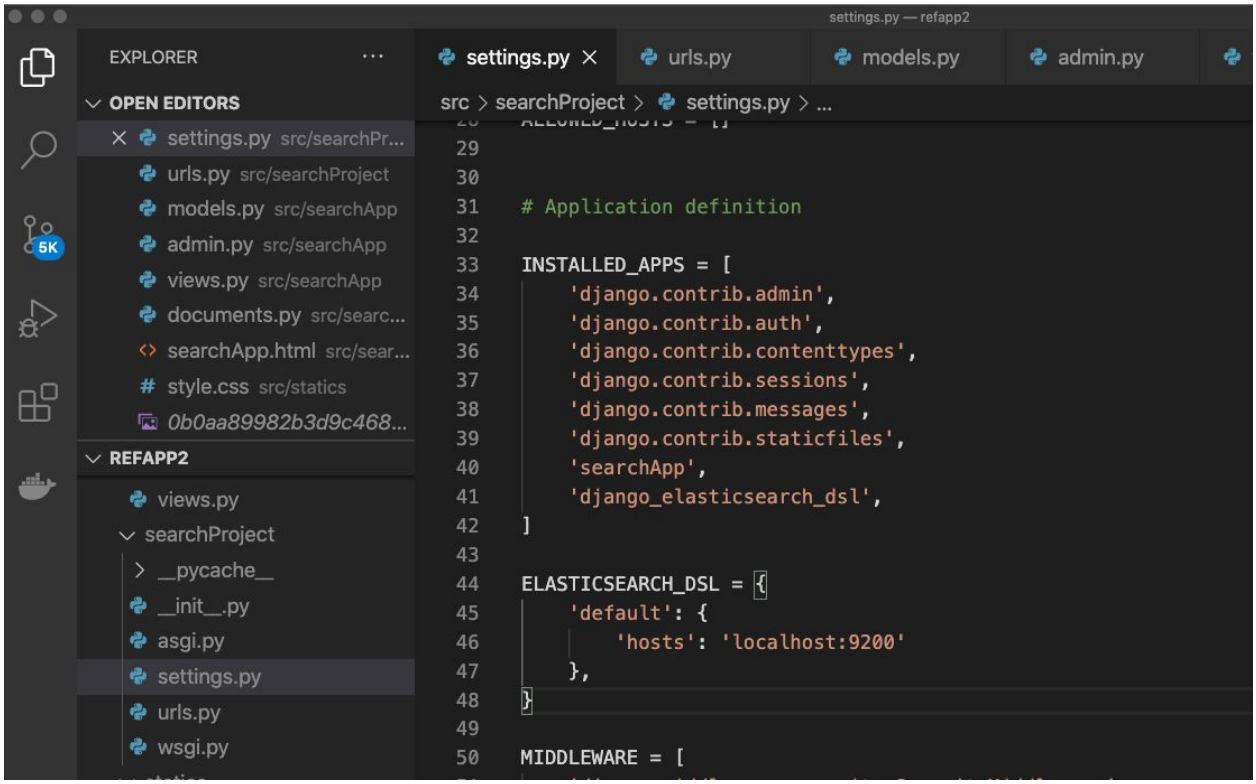
We firstly create the vector file for each image, and then use those vectors to generate the nearest_neighbors.json file.



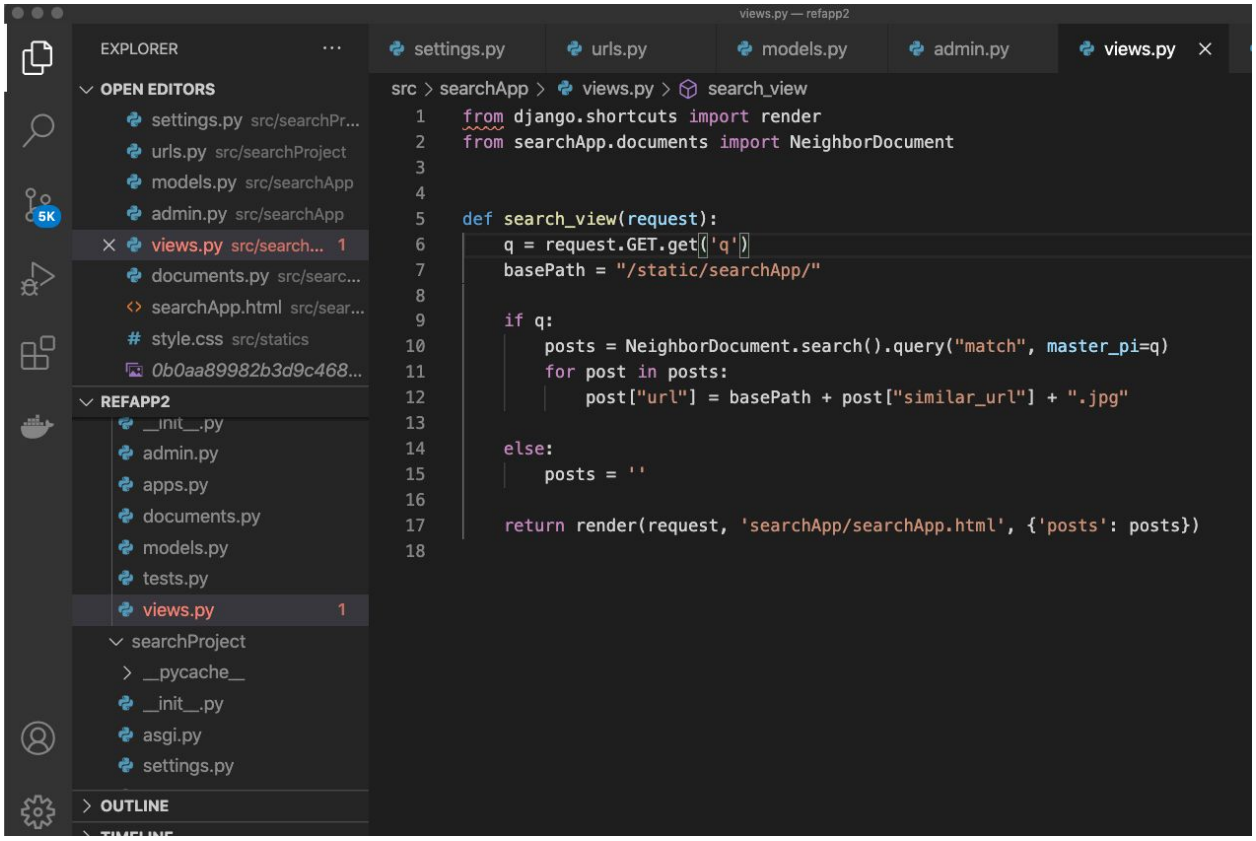


Integrate elasticsearch with django framework

We used **django-elasticsearch-dsl** for the integration. Run **python manage.py search_index --rebuild** to indexing the data model.



Search based on the N-Nearest Neighbor json file



Generate UI for the app