📖 docs_platform_take_home_test.MD

Loblaw Digital - Data Science Take-Home Task Instructions (ML SW Engineer Role)

Thank you for taking the time to complete our take-home task.

We know your time is valuable. While it is up to you to decide how long you want to work on this task, we do not expect you to spend more than 2 hours.

Please send us via email reply all relevant code, documentation, and analysis that you've produced, including elements we have not explicitly requested. Please clean up any cells/code/comments that do not add to your story.

You will also be required to submit your product recommendations electronically to our submission service as described below.

You may cite external resources such as Google and StackOverflow. You may use compute resources other than your personal computer if necessary, and if you do, please detail them in your documentation.

We ask that you do NOT share this task or the associated datasets with anyone.

Data

Data can be accessed through this link: `https://drive.google.com/drive/folders/107kFio90817OMypVsUxPIRSnjZFi-IF4`

Some details about the data are left vague on purpose.

- **`transactions.zip` (unzips to `transactions.txt`)**: each line is an ecommerce grocery order in json format. Together, they constitute 1.4 million transactions randomly sampled across a time period.
- **`products.txt`** : tab-separated product ID, MCH code (see below), and product name.
- **`mch_categories.tsv`** : Merchandise Hierarchy Category (MCH) Codes, which, as the name implies, constitute a hierarchy of product categories.
- **`ld-ds-take-home-service-account.json`** : a file containing a service account key that you will need to submit your response electronically (see below)

Task: product recommendation

In our product detail page view, we would like to show a `Recommended Products` component containing 5 recommended other products.

Please implement a baseline recommendation model based purely on *co-purchase frequency*. For example, if item A and item B were bought together in 5 separate transactions, their co-purchase frequency is 5.

Please write a function that takes a product ID and returns, using your model, the ID and name of the top 5 recommended products. Note that we need the serving time to be low for a good customer experience.

Please answer the following questions:

- What are the item ID and name of the top 5 co-purchased items for item with ID `20592676_EA` ? What about `20801754003_C15` ?
- How would you deploy and serve the model?

Please also submit your product recommendations to our test submission service using the service account key file included in the Google Drive. This service is deployed to Google App Engine and is protected by Google's Identity-Aware Proxy (IAP) (a common architecture in many real-world scenarios). This means you will need to:

1. Use the service account key file to obtain a `Credentials` object
2. Use the `Credentials` object to authenticate with the IAP
3. Send a POST request to `https://ld-ds-take-home-test.appspot.com/submissions` with a request body such as:

```
{
   'name': 'My Name',
   'email': 'myemail@somedomain.com',
   '20592676_EA': ['product1', 'product2', ... ],
   '20801754003_C15': ['product1', 'product2', ... ]
}
```

You can use the following code samples and resources to help you complete this section:

- https://cloud.google.com/bigquery/docs/authentication/service-account-file
- https://cloud.google.com/iap/docs/authentication-howto#iap_make_request-python
- https://cloud.google.com/iap/docs/concepts-overview