

CS 327E Project 4, due Thursday, 10/08.

This project makes use of the same shopify dataset which was used in Project 3. If you still have this data in your Jupyter notebook instance, you are good to go. If you don't have it anymore, you'll need to download as follows:

```
gsutil cp gs://cs327e-open-access/shopify.zip .
```

Redesign the shopify database for Firestore in accordance to Firestore's data model and the design principles discussed in class. Create an ERD with your Firestore database. The ERD should be created in Lucidchart. Be sure to include collections and embedded maps. Follow the college example from class for formatting and style. Collections should have document ids in addition to field names and types. Use one background color to identify collections and a different color to identify embedded maps. Draw the relationships between all entity types. If you're not sure what type of relationship exists between two given entities, sample the data. Export the ERD as a pdf file and name it `shopify-erd.pdf`.

Create a new jupyter notebook and name it `project4.ipynb`. Implement the following logic in your `project4.ipynb` notebook:

- Map entities to Firestore collections and embedded maps and populate them with the appropriate data from the shopify dataset.
- Get a document count for each collection.
- Implement the following queries:
 - 1) List the 10 apps with the highest number of reviews (based on `apps.review_count`). Return the `id`, `title`, `developer`, `rating` and `reviews_count` of those apps. Order the results by `reviews_count` in descending order.
 - 2) Find the apps which are categorized as "Trust and security". Return the `id`, `url`, `title`, `developer` and `rating` of those apps. Order the results by `rating` in descending order. Limit the number of results to 5.
 - 3) Find all the apps that are under the Premium Plan and whose price is \$12.99 or higher. Return the `id`, `title`, `developer`, `rating`, and `reviews_count` of those apps. Order the results by `price` in ascending order. Limit the number of results to 10.

CS 327E Project 4 Rubric

Due Date: 10/08/20

<p>Create an ERD for shopify designed for Firestore. Include field names, types, and ids. Draw proper relationships between entity types. -3 for each missing field, data type or id -3 for each missing or incorrect relationship between entity types</p>	20
<p>Create a notebook <code>project4.ipynb</code> -3 incorrect file name</p>	3
<p>Create a Firestore collection for each collection entity from the ERD -5 for each collection which does not match its entity specification -4 for each embedded map which does not match its entity specification -3 for each field which does not match its entity specification -2 for each id which does not match its entity specification</p>	15
<p>Populate each collection with its corresponding shopify data and get a count of the number of documents per collection. -5 for each empty collection -3 for each collection which has missing documents -3 for each missing or incorrect count</p>	20
<p>Implement Query #1: Apps with highest number of reviews. -3 incorrect or missing filter -3 incorrect or missing order by -3 incorrect or missing fields in results -3 incorrect number of results returned</p>	12
<p>Implement Query #2: Trust and security apps. -5 incorrect or missing join -3 incorrect or missing filter -3 incorrect or missing order by -3 incorrect or missing fields in results -3 incorrect number of results returned</p>	15
<p>Implement Query #3: Premium Plan apps. -5 incorrect or missing join -3 incorrect or missing filter -3 incorrect or missing order by -3 incorrect or missing fields in results -3 incorrect number of results returned</p>	15
<p><code>shopify-erd.pdf</code> pushed to your group's private repo on GitHub. Your project will not be graded without this submission.</p>	Required
<p><code>project4.ipynb</code> pushed to your group's private repo on GitHub. Your project will not be graded without this submission.</p>	Required

<p>submission.json submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema:</p> <pre>{ "commit-id": "your most recent commit ID from GitHub", "project-id": "your project ID from GCP" }</pre> <p>Example:</p> <pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	Required
Total Credit:	100