

CS 327E Project 5, due Thursday, 10/21.

This project makes use of the same Shopify dataset as [Project 3](#).

The goals of this project are to redesign the Shopify schema for Firestore, create the database objects according to your model, and populate the objects with the Shopify data.

Using Lucidchart, create an ERD of your Firestore schema. The schema should be modeled by following the design guidelines discussed in class and using as input the same access patterns as those used for [Practice Problem 1](#):

Access patterns:

1. Get apps by category (Category.title)
2. Get apps with highest review_count
3. Get pricing plan details by app (Apps.id)
4. Get key benefits by app (Apps.id)

Ensure that your diagram captures the field names and types for each Firestore collection and subcollection. Draw the appropriate relationships between collections. If you're unsure what type of relationship exists between two given collections, consult the Spanner schema for Shopify and/or sample the data. For readability, please use one background color to identify collections and a different color to identify subcollections. Follow the College example from class for other formatting and style guidance. Export your ERD as a pdf file and name it `shopify-firestore-erd.pdf`.

Create a new jupyter notebook and name it `project5.ipynb`. Implement the following logic in your `project5.ipynb` notebook:

- Download the dataset to your notebook instance:

```
gsutil cp gs://cs327e-open-access/shopify.zip .
```
- Create the Firestore collections and sub collections based on your design and populate them with the Shopify data.
- Get a document count for each collection and subcollection.
- List the top 10 "Productivity" apps (whose `categories.title = "Productivity"`) sorted by their rating in [descending order](#). Return the `id`, `title`, `developer`, `rating` and `reviews_count` for those apps. Limit the results to the first 10 records. Note: this is query refers to access pattern #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` for those apps. Order the results by `reviews_count` in [descending order](#).
Note: this is query refers to access pattern #2.

CS 327E Project 5 Rubric

Due Date: 10/21/21

<p>Create an ERD for the shopify data in Firestore. Include field names, data types, and ids for the collections and subcollections. Draw proper relationships between collections and subcollections.</p> <ul style="list-style-type: none"> -3 for each missing field name, data type or id -3 for each missing or incorrect relationship between entity types 	20
<p>Create the Firestore database objects that are represented in your ERD</p> <ul style="list-style-type: none"> -5 for each collection which does not match its entity specification -4 for each subcollection 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 	20
<p>Populate each collection with the appropriate shopify data and get a count of the number of documents per collection and subcollection.</p> <ul style="list-style-type: none"> -5 for each empty collection or subcollection -3 for each collection which has missing documents -3 for each missing or incorrect count 	30
<p>List the top 10 productivity apps with the highest rating.</p> <ul style="list-style-type: none"> -3 incorrect or missing filter -3 incorrect or missing order by -3 incorrect number of results returned -3 incorrect or missing fields in results 	15
<p>List the 10 apps with the highest number of reviews.</p> <ul style="list-style-type: none"> -3 incorrect or missing filter -3 incorrect or missing order by -3 incorrect or missing fields in results -3 incorrect number of results returned 	15
<p>shopify-firestore-erd.pdf pushed to your group's private repo on GitHub. Your project will not be graded without this submission.</p>	Required
<p>project4.ipynb 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 style="font-family: monospace;"> { "commit-id": "your most recent commit ID from GitHub", "project-id": "your project ID from GCP" } </pre> <p>Example:</p>	Required

<pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9", "project-id": "some-project-id" }</pre>	
Total Credit:	100