

# Data Challenge Documentation

Code ▼

## Instructions

- Explore the data included with these instructions and prepare a 45 minute long presentation with recommendations and insights about the App Store
- You may discuss one or more of the suggested topics below in your presentation but they are only suggestions; present whatever you judge to be the most valuable and actionable information from your analysis:
  - Mocking up a business dashboard for tracking the App Store's performance
  - Grouping accounts into meaningfully different segments and analyzing their behavior
  - Grouping apps into meaningfully different segments
- Just like with real data there may be incorrect, suspect, or missing records. If you find a data quality issue please note it in your presentation but analyze the data to the best of your abilities as though there were no issues

## Reference guide for tables

Data for this challenge comes with six data tables in CSV format.

- *account\_dat.csv* - metadata on accounts
- *app\_dat.csv* - metadata on apps
- *in-app\_dat.csv* - metadata on in-app content that can be purchased through a given app
- *device\_ref.csv* - metadata on devices
- *category\_ref.csv* - metadata on app categories (also called genres)
- *transaction\_dat.csv* - data on daily downloads and purchases of both apps and in-app content in the App Store

### **account\_dat.csv**

- *acct\_id* - a unique ID associated with an account
- *create\_dt* - date the account was created
- *payment\_type* - 'Free' for accounts with no payment method, 'PMOF' for accounts with a payment method on file

### **app\_dat.csv**

- *app\_name* - the name of the app
- *content\_id* - a unique ID associated with a piece of content; all the *content\_ids* in this table pertain to apps
- *category\_id* - an ID field that denotes the app's category (also called genre)
- *device\_id* - the device(s) where the app is available for purchase/download

### **in-app\_dat.csv**

- *parent\_app\_content\_id* - the *content\_id* of the app where the in-app content can be purchased
- *content\_id* - a unique ID associated with a piece of content; all the *content\_ids* in this table pertain to

purchaseable in-app content

- type - the type of in-app content

#### **device\_ref.csv**

- device\_name - the name of the device
- device\_id - an ID field that denotes the type of device

#### **category\_ref.csv**

- category\_name - the name of the category (also called genre)
- category\_id - an ID field that denotes the category of an app

#### **transaction\_dat.csv**

- create\_dt - the date a transaction occurred
- content\_id - the content\_id of the item that was purchased/downloaded
- acct\_id - the id of the account making the purchase/download
- price - the price paid in USD
- device\_id - the type of device where the content was downloaded/purchased