

CSC4008 - Assignment 1

Yohandi

October 13, 2023

Sales Analysis

The total sales for the 3 types

```
+-----+-----+
|Type|          Total Sales|
+-----+-----+
|  A | 4.331014722750015E9|
|  B | 2.0007007368200486E9|
|  C | 4.055035275399986E8|
+-----+-----+
```

The average sales on Holidays vs. Non-Holidays

```
+-----+-----+
|IsHoliday|      Average Sales|
+-----+-----+
|   false |15901.445069008514|
|    true | 17035.82318735042|
+-----+-----+
```

Sales are generally higher during holidays.

People You Might Know

Spark pipeline sketch

First, load the `ego-facebook.txt` into a Spark DataFrame, treating each friendship as a bidirectional edge. Perform a self-join on the data to count mutual connections but exclude direct friendships and self-friendships. For each user, sort the recommendations based on mutual friend counts and user ID (as a secondary sort).

Recommendations for the users with the following user IDs: 10, 152, 288, 603, 714, 1525, 2434, 2681

```
Recommendations for 10: [2, 3, 4, 5, 6, 7, 8, 9, 11, 12]
Recommendations for 152: [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
Recommendations for 288: [71, 1525, 69, 90, 217, 2348, 2351, 2352, 2354, 2356]
Recommendations for 603: [1, 289, 290, 291, 292, 293, 294, 295, 296, 297]
Recommendations for 714: [1, 712, 713, 715, 717, 718, 1525, 90, 217, 247]
Recommendations for 1525: [288, 1, 710, 714, 603]
Recommendations for 2434: [71, 288, 711, 716, 719, 720, 2348, 2351, 2352, 2354]
Recommendations for 2681: [71, 288, 710, 711, 716, 719, 720, 721, 722, 2348]
```