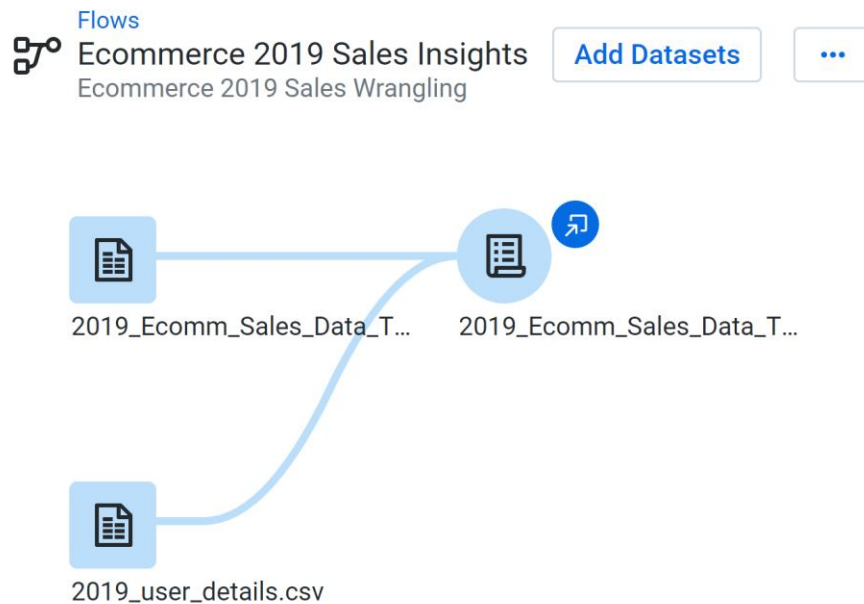


Trifacta Wrangler –

1. Created a flow and imported the datasets. Joined them.



2. Created Recipe for Transformations –

Ecomm_behavior_Sales_Trifacta
Initial Sample

Run Job

#	column2	
🕒	Event_date	
ABC	DayofWeek	
🕒	Event_time	
#	Event_hour	
ABC	Event_timezone	
ABC	event_type	
#	product_id	
#	category_id	
ABC	Category	
ABC	Sub_category1	
ABC	Sub_category2	

16 Columns 45,055 Rows 5 Data Types

Details

ABC Sub_category1 ...

Unique Values

Q Search...

smartphone	16,483
kitchen	5,784
audio	2,712
environment	2,128
clocks	1,901
accessories	1,706
video	1,649
tools	1,639
notebook	1,611
shoes	1,513
living_room	834
bicycle	713
components	615
bedroom	591

Used If condition for creating Day of Week column.

Group by to get total sales per brand –

ECOMMERCE BEHAVIOUR SALES INSIGHTS >
Ecomm_behavior_Sales_Trifacta - 2
 Initial Sample

Run Job

New Step Recipe

ABC	brand	ABC	DayofWeek	##	sum_price
925 Categories		7 Categories		1 - 836,200	
Huawei1		Monday			38676.13
Lg		Thursday			49725.710000000003
Saturn		Saturday			510.0400000000001
Samsung		Thursday			352326.25000000003
Cenmax		Tuesday			210.09999999999997
Redmond		Saturday			3810.1900000000005
Xiaomi		Saturday			104065.49999999999
Oris		Saturday			4729.86
Panasonic		Tuesday			1996.9999999999998
Samsung		Tuesday			386854.88000000007
Salamander		Wednesday			1076.73
Sony		Tuesday			53361.9599999999985
Midea		Saturday			4047.8300000000001
Apple		Tuesday			782542.42999999989
Axis		Saturday			386.11
Ho		Sunday			31406.449999999999

3 Columns 3,619 Rows 2 Data Types

1 Create sum_price from SUM(price) grouped by brand, DayofWeek

Full Recipe created –

New Step Recipe

1 Rename column2 to 'Index'

2 Convert text in 3 columns to Propercase

3 Split event_time on delimiters matching '' into 2 columns

4 Rename event_time1 to 'event_date'

5 Split event_time2 on delimiters matching '' into 2 columns

6 Rename event_time1 to 'event_time'

7 Rename event_time3 to 'event_timezone'

8 Create event_hour from HOUR(event_time)

9 Split category_code on delimiters matching '' into 2 columns

New Step

Recipe

×

□ ...

⚙

- Set sub_category2 to `IFMISSING($col, sub_category1)`
- Change date format of event_date to yyyy-MM-dd
- Create DayofWeek from `WEEKDAY(event_date)`
- Create Day_of_Week from `IF(DayofWeek == 1, 'Monday', IF(DayofWeek == 2, 'Tuesday', IF(DayofWeek == 3, 'Wednesday', IF(DayofWeek == 4, 'Thursday', IF(DayofWeek == 5, 'Friday', IF(DayofWeek == 6, 'Saturday', 'Sunday')))))`
- Delete DayofWeek
- Left join with 2019_user_details.csv on `user_id == user_id`

Published the Final Dataset to be used in Salesforce –

Ecommerce 2019 Sales Insights > 2019_Ecomm_Sales_Data_Trifacta

Job 125230

Finished Today at 1:57 AM

Download results

...

Overview

Output Destinations

Profile

Dependencies

✓ Transform

Completed Today at 1:56 AM, started Today at 1:53 AM • Ran for 3 min

Environment Spark

View steps and dependencies

✓ Profile

Completed Today at 1:57 AM, started Today at 1:56 AM • Ran for 2 min

100% valid values

0% mismatching values

0% missing values

View profile

✓ Publish

Completed Today at 1:56 AM, started Today at 1:56 AM • Ran for <1 sec

Activity

2019_Ecomm_Sales_Data_Trifacta.csv

Completed

Job summary

Job ID

125230

Job status

Completed

Flow

Ecommerce 2019 Sales Insights

Output

2019_Ecomm_Sales_Data_Trif...

Execution summary

Job type

Manual

User

Nikhil Kohli

Start time

February 12th 2020, 1:53 am

Finish time

February 12th 2020, 1:57 am

Last update

February 12th 2020, 1:57 am

Duration

4 minutes