## databricks02 - Manage Query Results



### **Create Tables**

Run the cell below to create tables for the questions in this notebook.

%run ../Utilities/02-CreateTables

Declared the following table:

products

Declared the following table:

raceResults

# **Question 1: Dedupe Sort**

## Summary

Deduplicate and sort data on products that need to be restocked in a grocery store.

### Steps to complete

Write a SQL query on the **products** table that achieves the following:

Removes duplicate rows

- Sorts rows by the aisle column in ascending order (with nulls appearing last), and then by **price** in ascending order
- Stores the results into a temporary view named q1Results

A properly completed solution should return both a DataFrame that looks similar to this:

itemId	amount	aisle	price
9958	64	2	2
1432	23	3	24
3242	14	5	5
7064	34	null	24

#### CREATE OR REPLACE TEMPORARY VIEW q1Results AS **SELECT**

**FROM** 

(SELECT

**DISTINCT** itemId,

amount,

aisle,

price

**FROM** products

**SORT BY** aisle **ASC**)

**SORT BY** price **ASC** 

OK

**SELECT** \* **FROM** q1Results

	itemId 📤	amount 🔺	aisle 🔺	price 📤
1	9958	64	2	2
2	6693	17	14	3
3	3242	14	5	5
4	9382	15	7	8
5	6914	5	null	11

6	7012	129	10	11
7	1432	23	3	24

Showing all 9 rows.

# **Question 2: Limit Results**

## Summary

Return the top five results for data that matches a set of criteria.

## Steps to complete

Write a SQL query on the table | raceResults | that achieves the following:

- Casts lastFinish column as date and renames to raceDate
- Sorts rows by the winodds column in descending order
- Limits the results to the top 5 winodds
- Stores the results into a temporary view named q2Results

A properly completed solution should return a DataFrame that looks similar to this:

name	winOdds	raceDate
Dolor Incididunt	.9634252	2015-07-29
Excepteur Mollit	.9524401	2019-08-15
Magna Ad	.9420442	2017-05-12
Sed In	.9325211	2014-08-29
Qui Cupidat	.9242451	2011-08-23

#### CREATE OR REPLACE TEMPORARY VIEW q2Results AS **SELECT**

name,

winOdds,

CAST(lastFinish AS date) raceDate

**FROM** 

raceResults

SORT BY winOdds DESC

LIMIT 5

OK

**SELECT** \* **FROM** q2Results

	name	winOdds 🔺	raceDate 🔺
1	Mollit Minim	0.99992424	2021-06-17
2	Adipiscing In	0.99943733	2002-01-03
3	Aliqua Sed	0.99885786	2001-11-07
4	Anim Deserunt	0.9984048	2019-01-23
5	Laboris Incididunt	0.99732566	2017-02-07

Showing all 5 rows.

© 2020 Databricks, Inc. All rights reserved.

Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation (http://www.apache.org/).