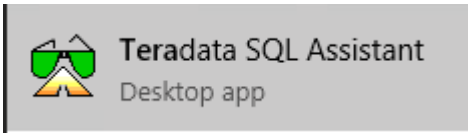
Dillards 2016 Data Extraction

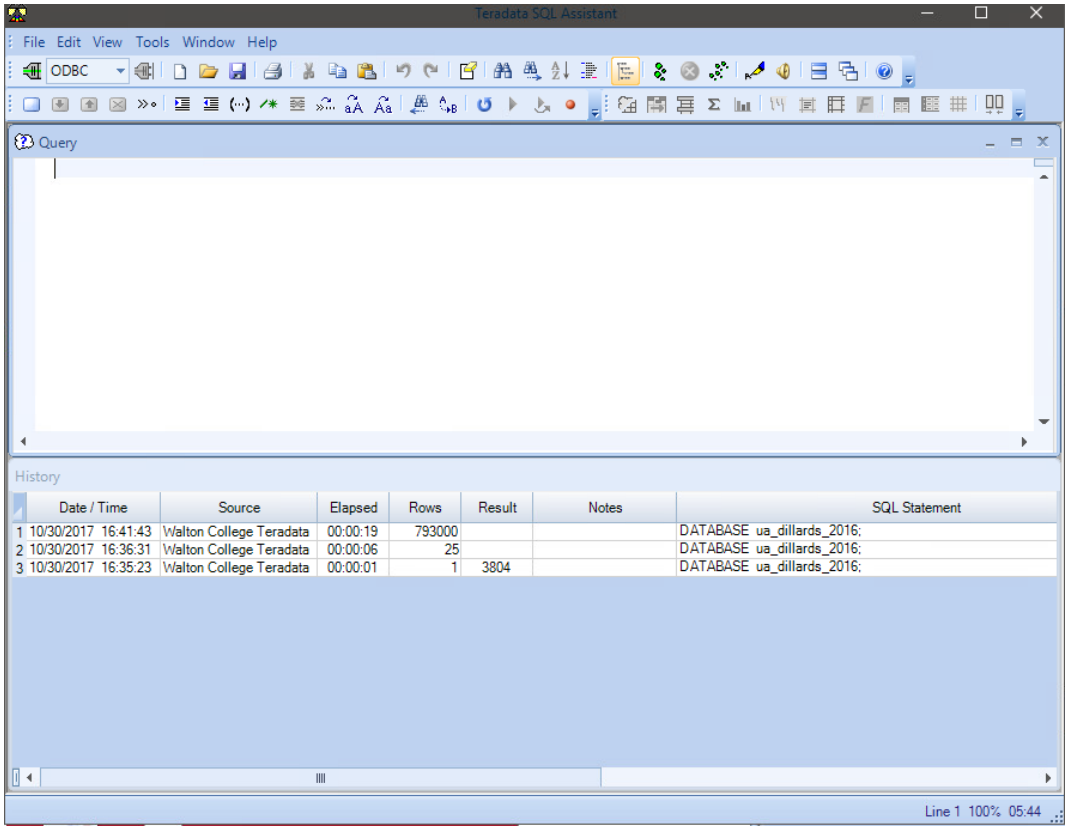
Teradata Connection

This tutorial assumes you have access to the University of Arkansas VMWare client. Note: your request for access needs to include permissions to create a VIEW in TUN. The intent of this exercise is to extract data from the Dillard's 2016 dataset into an Excel file in order to load the information into an infocube or the application of your choice.

NOTE: the data used during this workshop should not be downloaded to your personal drives and should remain on the Remote Desktop S: drive provided by the University of Arkansas. This is due to our agreement with the data providers.

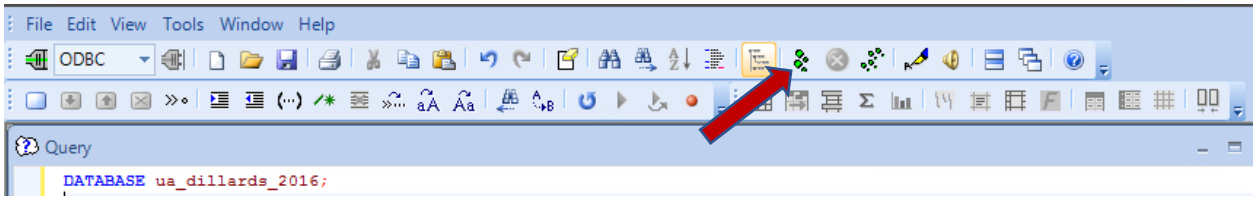
1. From the Desktop or Start Menu, search for and open the software Teradata SQL Assistant. You will see a Windows similar to the below screenshot.





2. Copy and paste the query below onto the Query window to access the UA_Dillards_2016 and click on the "footsteps" execute button

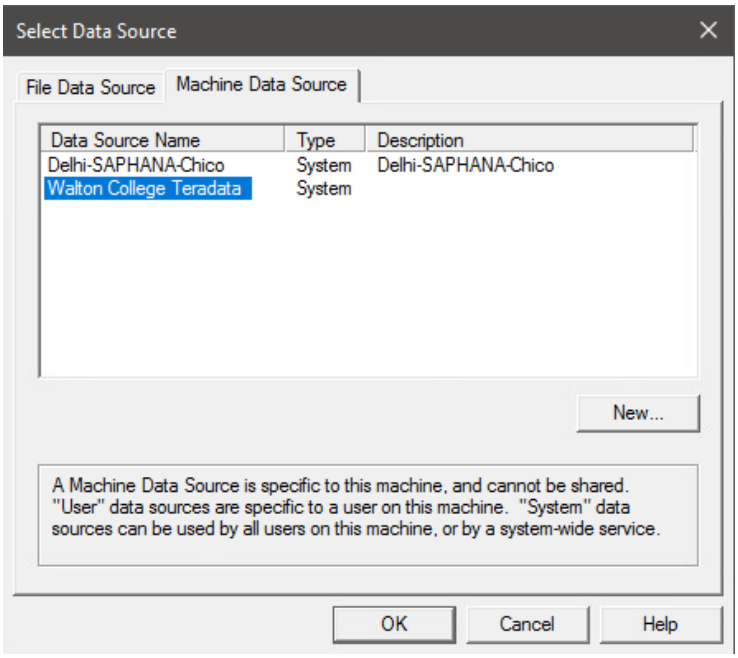
```
DATABASE ua_dillards_2016;
```



Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

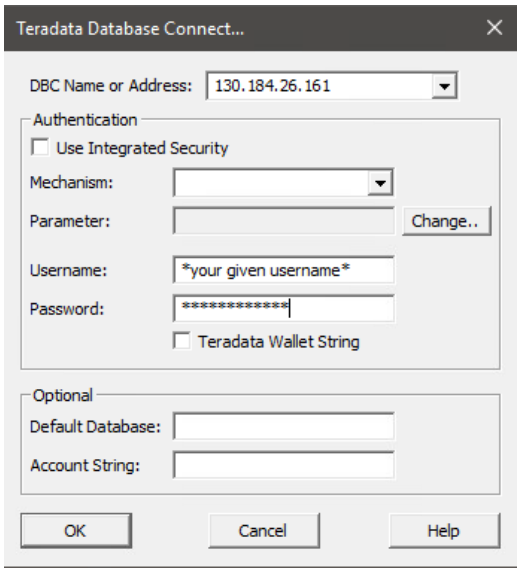
Dillards 2016 Data Extraction

3. A Select Data Source window will pop up to ask you the credentials to access the Dillard's database. Navigate to the **Machine Data Source** tab on top. Then double-click on **Walton College Teradata** (as shown to the right).



4. Click **OK**

5. Here you will see another pop-up window called **Teradata Database Connect…**



6. Enter the **Username**: and **Password**: provided by your instructor

7. Click **OK**

8. Congratulations! Now you are connected to the UA_Dillards_2016 database. You are ready to extract the data needed for your assignment.

SQL Query to extract data

9. We will need to extract data from three different tables. Since the Dillard's 2016 dataset has approximately 450 million rows, we will need to narrow the data extraction down. To do this, we will run three queries. On each of these queries, you will run the provided SQL code, delete that code and then move to the next query. The first two queries create a view with the **CREATE VIEW** statement. This is a way to create a temporary sub table, without interfering with the real database. For each view, there will be a filter to extract part of the data. The third query will use the two views and a third table to join all of the extracted data. You should have 510,161 records in the final selection.

Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

Dillards 2016 Data Extraction

10. **CREATE VIEW** – Tran
    a. Navigate to the query window and delete any current code
    b. This view extracts data from the TRANSACT table only
    c. The filter for this view is the STORE code between 400 and 450
    d. Copy and paste this SQL query (below) into the query window

```
CREATE VIEW Tran AS
SELECT TRANSACTION_ID, STORE, ITEM_ID, TRAN_DATE, SKU, TRAN_TYPE, ORIG_PRICE,
TRAN_AMT
FROM TRANSACT
WHERE STORE BETWEEN 400 and 450;
```

    e. Click the "footsteps" execute button  on the menu bar.

11. **CREATE VIEW** – Skus
    a. Navigate to the query window and delete any current code
    b. This view extracts data from the SKU_STORE table only
    c. The filter for this view is the STORE between 400 and 450
    d. Copy and paste this SQL query (below) into the query window

```
CREATE VIEW Skus AS
SELECT STORE, COST, SKU, RETAIL
FROM SKU_STORE
WHERE STORE BETWEEN 400 and 450;
```

    e. Click the "footsteps" execute button  on the menu bar.

12. **SELECT** – Tran
    a. Navigate to the query window and delete any current code
    b. This view extracts data from the SKU_STORE table only
    c. The filter for this table is the ST.STATE of Arkansas
    d. There is also a filter to eliminate brand.name having the special characters of ?, ., and &.
    e. Copy and paste this SQL query (below) into the query window

```
SELECT DISTINCT T.transaction_id, T.tran_date, T.tran_type, T.orig_price,
T.tran_amt,
        S.item_id, S.dept, S.brand_name,
        ST.store, ST.city, ST.state, ST.zip_code,
        SKS.cost
FROM tran T
    JOIN sku S ON T.item_id = S.item_id
    JOIN store ST ON T.store = ST.store
    JOIN skus SKS ON T.sku = SKS.sku
WHERE (ST.state = 'AR') AND NOT(S.brand_name = '?') AND NOT(S.brand_name LIKE
'%+%') AND NOT(S.brand_name LIKE '%''%') AND NOT(S.brand_name LIKE '%=%') AND
NOT(S.brand_name LIKE '%-%')AND NOT(S.brand_name LIKE '%.%') AND
NOT(S.brand_name LIKE '%&%') AND NOT(S.brand_name LIKE '%/%')AND
NOT(S.brand_name LIKE '%180s%') AND NOT(S.brand_name LIKE '%1st%') AND
NOT(S.brand_name LIKE '%2 Hip%')AND NOT(S.brand_name LIKE '%7 for%');
```

Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

Dillards 2016 Data Extraction

13. Click the “footsteps” execute button  on the menu bar.

14. The final query should take **approximately 30 seconds to 1 minute to run** depending on the connection from your computer. You should have 510,161 rows of records in the final selection. Your result table should look like the picture below and the records can be verified on this screen.

Query (Walton College Teradata)

```
SELECT DISTINCT T.transaction_id, T.tran_date, T.tran_type, T.orig_price, T.tran_amt,
    S.item_id, S.dept, S.brand_name,
        ST.store, ST.city, ST.state, ST.zip_code,
        SKS.cost
FROM tran T
    JOIN sku S ON T.item_id = S.item_id
    JOIN store ST ON T.store = ST.store
    JOIN skus SKS ON T.sku = SKS.sku
WHERE (ST.state = 'AR') AND NOT(S.brand_name = '?') AND NOT(S.brand_name LIKE '%+%') AND NOT(S.brand_name LIKE '%''%') AND
```

Answerset 1

	TRANSACTION_ID	TRAN_DATE	TRAN_TYPE	ORIG_PRICE	TRAN_AMT	ITEM_ID	DEPT	BRAND_NAME	STORE	CITY	STATE	ZIP_CODE	COST
1	56,822,355	7/20/2015	P	68.00	44.90	12,150,000	227	Natori	403	FAYETTEVILLE	AR	72703	1.00
2	100,154,479	8/24/2016	P	28.00	5.88	47,158,237	239	Sleep Sense	406	JONESBORO	AR	72401	1.00
3	93,002,968	5/29/2016	P	28.00	28.00	39,487,998	481	Gossip Girl	405	LITTLE ROCK	AR	72205	1.00
4	92,982,411	6/8/2016	P	108.00	108.00	39,248,064	198	Miraclesuit	413	FORT SMITH	AR	72903	1.00
5	91,702,099	4/29/2016	P	149.00	36.50	26,909,775	133	Jessica Howard	403	FAYETTEVILLE	AR	72703	1.00
6	88,151,137	4/21/2016	P	129.00	77.40	36,968,341	189	Alex Marie	407	NORTH LITTLE ROCK	AR	72116	1.00
7	77,146,827	1/9/2016	P	169.00	101.40	31,468,298	161	Antonio Melani	405	LITTLE ROCK	AR	72205	1.00
8	88,700,477	3/26/2016	P	318.00	318.00	17,448,295	618	Frye	408	ROGERS	AR	72758	1.00
9	20,959,038	7/24/2014	P	9.00	9.00	12,365,984	166	Sugarlips	406	JONESBORO	AR	72401	1.00
10	83,178,159	2/27/2016	P	158.00	47.40	18,458,104	618	Frye	405	LITTLE ROCK	AR	72205	1.00
11	41,749,268	2/3/2015	P	59.00	59.00	15,863,688	181	Calvin Klein	406	JONESBORO	AR	72401	1.00
12	81,771,042	2/27/2016	P	39.00	17.55	41,497,942	166	RD Style	413	FORT SMITH	AR	72903	1.00
13	94,383,307	6/25/2016	P	35.00	35.00	30,048,115	157	Lauren Ralph Lauren	408	ROGERS	AR	72758	1.00

History

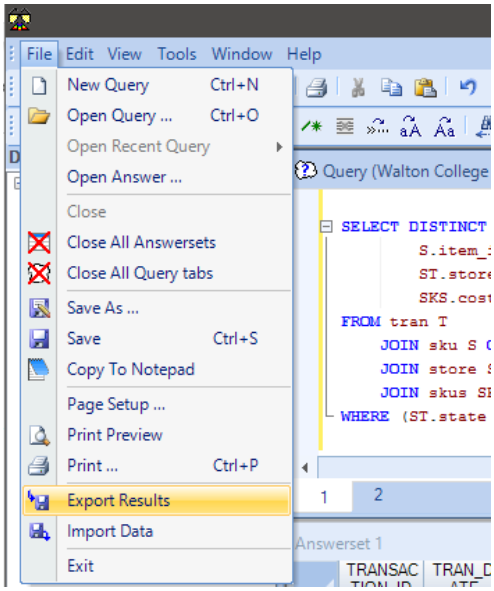
	Date / Time	Source	Elapsed	Rows	Result	Notes	SQL Statement	Length
1	5/14/2018 3	Walton Coll	00:01:05	510161			SELECT DISTINCT T.transaction_id, T.tran_date, T.tran_type, T.orig_price, T.t	750
2	5/14/2018 (	Walton Coll	00:00:00	0	3804		CREATE VIEW Skus AS	102
3	5/14/2018 (	Walton Coll	00:00:00	0	3804		CREATE VIEW Tran AS	156
4	5/14/2018 (	Walton Coll	00:00:00	1			DATABASE ua_dillards_2016;	26

Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

Dillards 2016 Data Extraction

Export result table in a text file

Now we will export to a file in order to provide access to the application of our choice



15. Navigate to File on the top bar, and select Export Results…

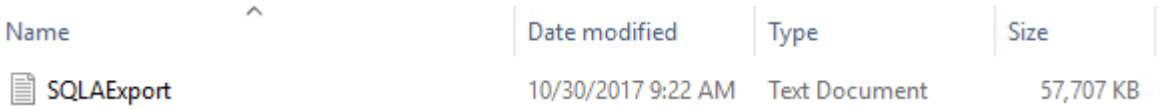
16. Now click the "footsteps"  execute button again to have Teradata run the query again and export the result directly.

17. You will now see a pop up window, prompting you for the location you want to save

18. Select your desired location, the File name: and make sure you Save as type: Delimited Text [ANSI] (*.txt).

19. **IMPORTANT**: after you click Save, the Teradata will take a while to export everything. It will take about 3 minute or less depends on your computer, since there are 510,161 rows of data to be export.

20. After MS SQL Server Studio finishes with saving the result table as txt file in your folder, go to the location that you saved it to make sure it exports successfully.



Import Text file into Excel

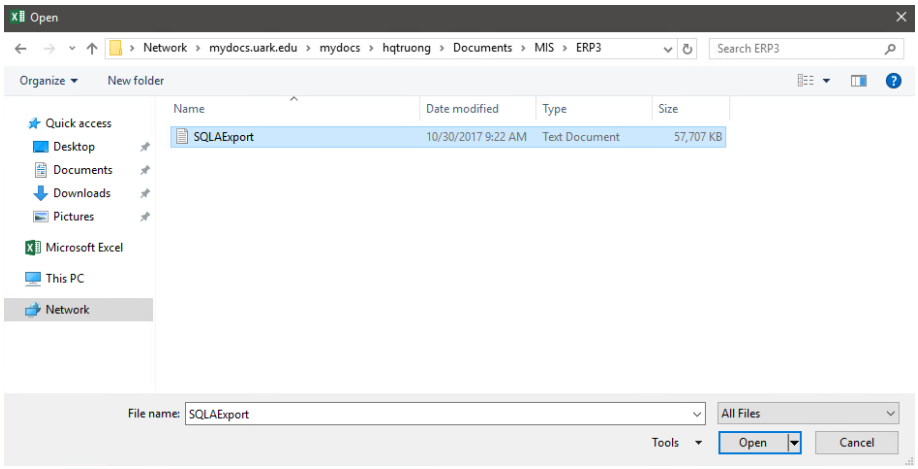
21. Open MS Excel.

22. Select **File** -> **Open** and Navigate to your .txt file folder

23. Choose the file format as **All Files** (as shown below)

24. Select your file

25. Click **Open**.



Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

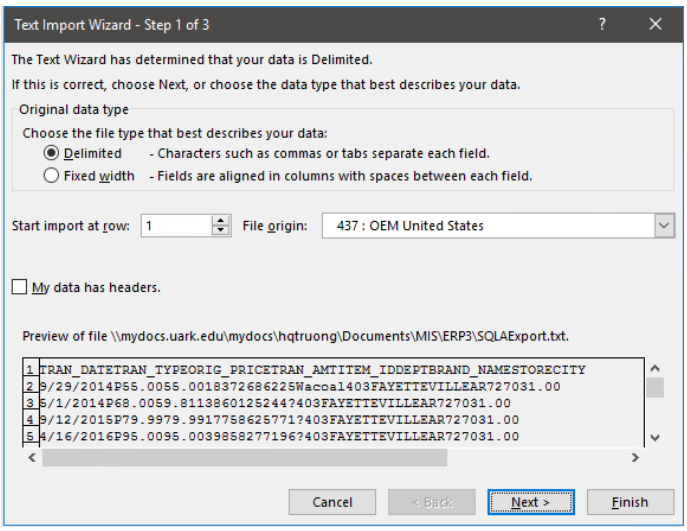
Dillards 2016 Data Extraction

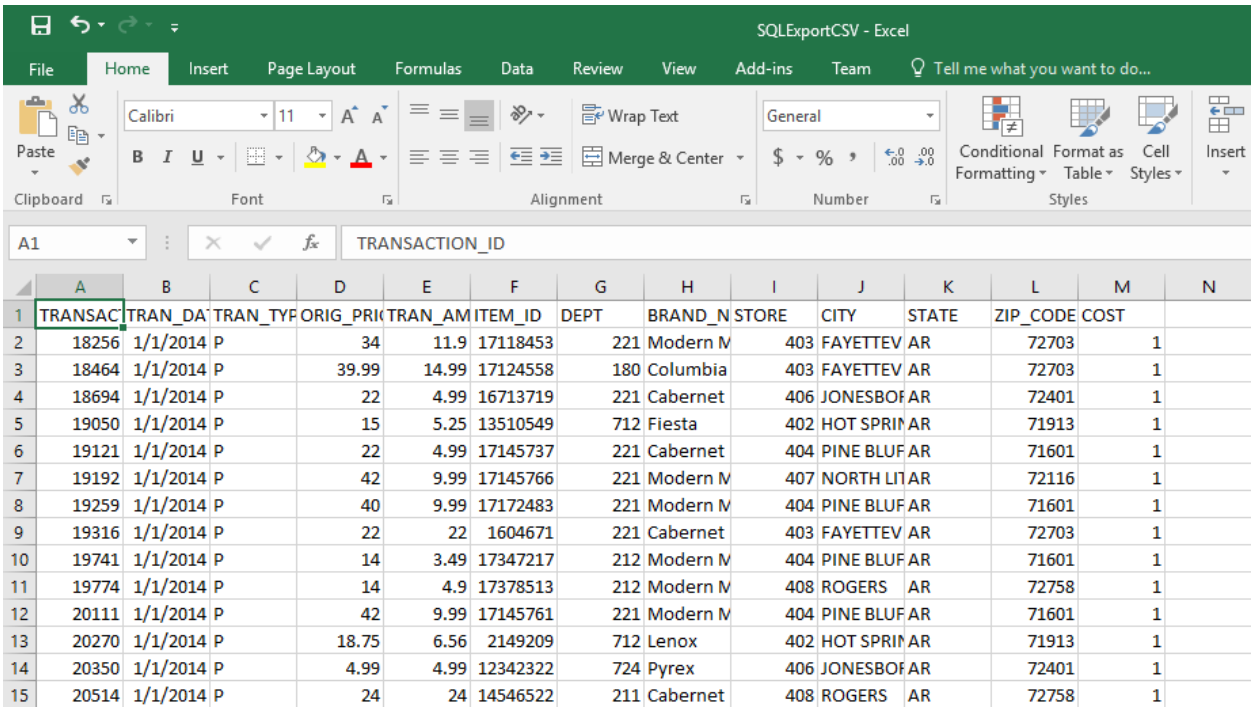
26. You will now see a window called **Text Import Wizard**, with different options of how you want to format your Excel file from a text file.

27. Accept the default settings and click **Finish** to import the file.

28. After couple seconds, you will now see the result is successfully imported into Excel. See next page for a screen shot.

29. **Save** the Excel file after you are done reviewing the data as a .csv comma delimited excel file





	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TRANSAC	TRAN_DA	TRAN_TYP	ORIG_PRI	TRAN_AM	ITEM_ID	DEPT	BRAND_N	STORE	CITY	STATE	ZIP_CODE	COST	
2	18256	1/1/2014	P	34	11.9	17118453	221	Modern M	403	FAYETTEV	AR	72703	1	
3	18464	1/1/2014	P	39.99	14.99	17124558	180	Columbia	403	FAYETTEV	AR	72703	1	
4	18694	1/1/2014	P	22	4.99	16713719	221	Cabernet	406	JONESBOF	AR	72401	1	
5	19050	1/1/2014	P	15	5.25	13510549	712	Fiesta	402	HOT SPRI	AR	71913	1	
6	19121	1/1/2014	P	22	4.99	17145737	221	Cabernet	404	PINE BLUF	AR	71601	1	
7	19192	1/1/2014	P	42	9.99	17145766	221	Modern M	407	NORTH LI	AR	72116	1	
8	19259	1/1/2014	P	40	9.99	17172483	221	Modern M	404	PINE BLUF	AR	71601	1	
9	19316	1/1/2014	P	22	22	1604671	221	Cabernet	403	FAYETTEV	AR	72703	1	
10	19741	1/1/2014	P	14	3.49	17347217	212	Modern M	404	PINE BLUF	AR	71601	1	
11	19774	1/1/2014	P	14	4.9	17378513	212	Modern M	408	ROGERS	AR	72758	1	
12	20111	1/1/2014	P	42	9.99	17145761	221	Modern M	404	PINE BLUF	AR	71601	1	
13	20270	1/1/2014	P	18.75	6.56	2149209	712	Lenox	402	HOT SPRI	AR	71913	1	
14	20350	1/1/2014	P	4.99	4.99	12342322	724	Pyrex	406	JONESBOF	AR	72401	1	
15	20514	1/1/2014	P	24	24	14546522	211	Cabernet	408	ROGERS	AR	72758	1	

30. There are three items to modify in your excel file prior to importing into your infocube.

    a. Replace all spaces in the Brand column with an underscore (_). 400,263 replacements

    b. Replace all spaces in the City column with an underscore (_). 389,612 replacements

    c. Add a column on the far right that says CURRENCY. Fill all cells with USD.

Created 2/28/2018 by Huy Truong & Ron Freeze
Updated 5/14/2018 by Ron Freeze

Dillards 2016 Data Extraction

31. You are now ready to import the data into your application.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TRANSACTION_ID	TRAN_DATE	TRAN_TYPE	ORIG_PRICE	TRAN_AMT	ITEM_ID	DEPT	BRAND_NAME	STORE	CITY	STATE	ZIP_CODE	COST	CURRENCY
2	18256	1/1/2014	P	34	11.9	17118453	221	Modern_Movement	403	FAYETTEVILLE	AR	72703	1	USD
3	18464	1/1/2014	P	39.99	14.99	17124558	180	Columbia	403	FAYETTEVILLE	AR	72703	1	USD
4	18694	1/1/2014	P	22	4.99	16713719	221	Cabernet	406	JONESBORO	AR	72401	1	USD
5	19050	1/1/2014	P	15	5.25	13510549	712	Fiesta	402	HOT_SPRINGS	AR	71913	1	USD
6	19121	1/1/2014	P	22	4.99	17145737	221	Cabernet	404	PINE_BLUFF	AR	71601	1	USD
7	19192	1/1/2014	P	42	9.99	17145766	221	Modern_Movement	407	NORTH_LITTLE_ROCK	AR	72116	1	USD
8	19259	1/1/2014	P	40	9.99	17172483	221	Modern_Movement	404	PINE_BLUFF	AR	71601	1	USD
9	19316	1/1/2014	P	22	22	1604671	221	Cabernet	403	FAYETTEVILLE	AR	72703	1	USD
10	19741	1/1/2014	P	14	3.49	17347217	212	Modern_Movement	404	PINE_BLUFF	AR	71601	1	USD
11	19774	1/1/2014	P	14	4.9	17378513	212	Modern_Movement	408	ROGERS	AR	72758	1	USD
12	20111	1/1/2014	P	42	9.99	17145761	221	Modern_Movement	404	PINE_BLUFF	AR	71601	1	USD
13	20270	1/1/2014	P	18.75	6.56	2149209	712	Lenox	402	HOT_SPRINGS	AR	71913	1	USD
14	20350	1/1/2014	P	4.99	4.99	12342322	724	Pyrex	406	JONESBORO	AR	72401	1	USD