ITP 249 Final Project Answer Sheet

Group Number:

_____17_____

Group Member Names:

___Verena Girgis, Carson Levy, Brianna Sun___

Question 1: What is the primary key here?

Customer

**Question 2: Insert a screenshot of a crosstab that shows Marital_Status vs Total_Discount. Display Totals.**

Exercise 4	Exercise 4_ ⋮			
	Marital_Status ▲	Married	Single	Total
Total_Discount	▲	Frequency	Frequency	Frequency
0%		73	24	97
5%		20	3	23
10%		151	28	179
14%		1	1	2
15%		161	39	200
19%		24	5	29
20%		313	90	403
24%		57	10	67
Total		800	200	1,000

Question 3: Insert a screenshot of the histogram of age with 15 bins.
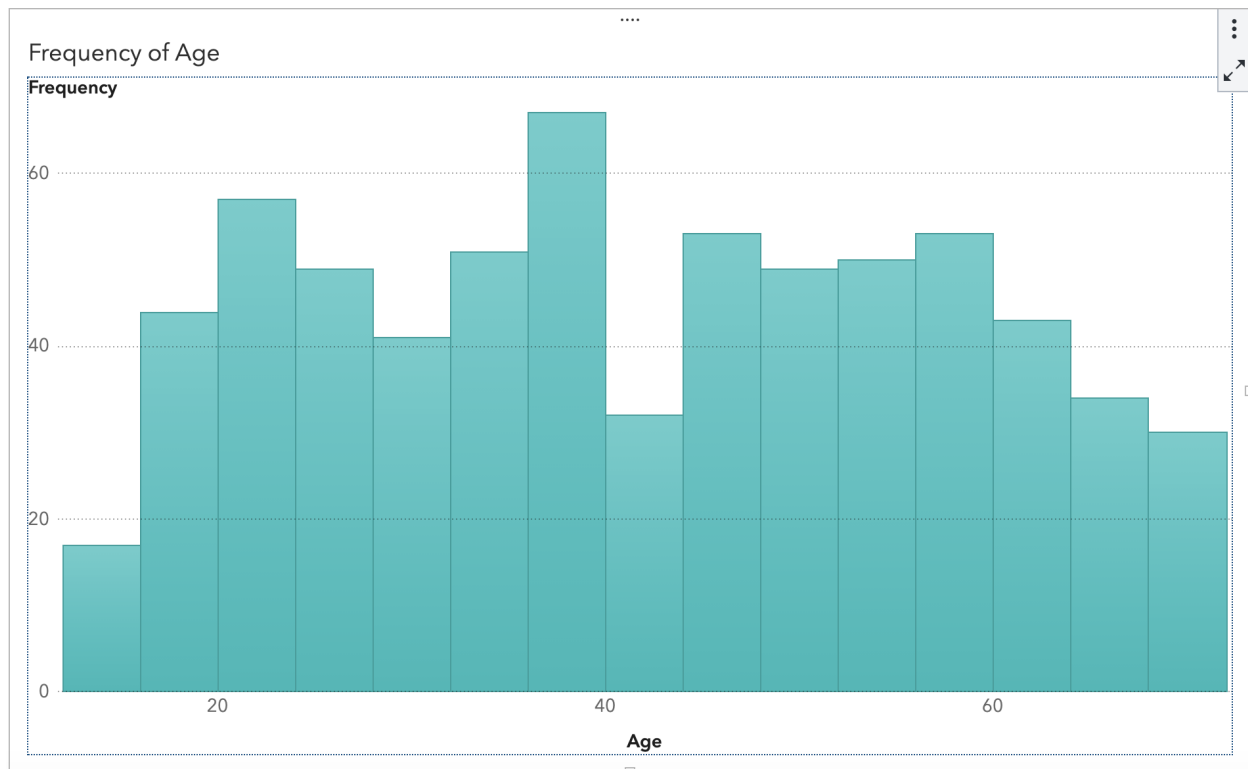
Question 4: How many variables (and names) are being displayed in your histogram?

Two Variables are being displayed - Type of Customer (Promotional) and Age

Question 5: Find the number of rows in the razorback stores data.

1000

Question 6: Which pair of variables is most correlated?
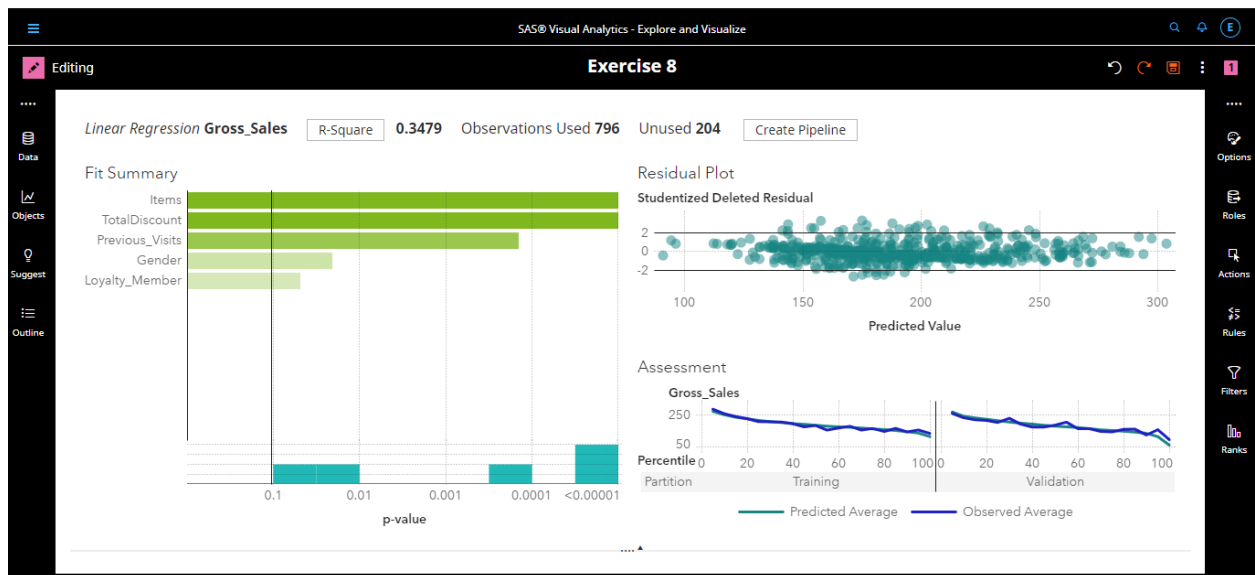
(Gross_Sales, Net_Sales)

Question 7: Which pair of variables is least correlated?

(Age, Previous_Visits)

Question 8: Insert a screenshot of your page

Question 9: Insert a screenshot of your Page for Exercise 8

Question 10: Which variables are most significant in your linear regression of the Gross_sales? Why?

Items and total discount; If the p-value is smaller, it means it is more significant

Question 11: Find the number of rows in the train partition and the validation partition.

Number of rows in train partition: 567

Number of rows in validation partition: 229

Question 12: What is a confusion matrix?

A confusion matrix displays the number predicted correct against the number that should be correct with the number predicted incorrect against the number incorrect

Question 13: Insert a screenshot of the confusion matrix.

Predicted	Observed	Training Frequency	Training Percentage	Validation Frequency	Validation Percentage
0	0	3,121	98.99%	1,352	98.90%
1	0	32	1.01%	15	1.10%
0	1	108	31.12%	39	29.32%
1	1	239	68.88%	94	70.68%

Question 14: How many observations were used for validation? Of those how many did the model misclassify?

1352 + 15 + 39 + 94 = 1731 observations, or 30% of the total dataset. The two dichotomies for predicted and observed are misclassified 15 + 39.

Predicted	Observed	Training Frequency	Training Percentage	Validation Frequency	Validation Percentage
0	0	3,121	98.99%	1,352	98.90%
1	0	32	1.01%	15	1.10%
0	1	108	31.12%	39	29.32%
1	1	239	68.88%	94	70.68%

Question 15: List the predictor variables from most important to least important.

Education, Income, Family, CCAvg, CD Account, Experience, Credit Card, Mortgage, Online, Securities Account

Question 16: How many observations were used for validation? Of those how many did the model misclassify? Insert a screenshot to support your answer.

1500. 28 were misclassified.

Node Statistics	Node Rules	Variable Importance	Cost-Complexity Pruning	Confusion Matrix	Lift	ROC	Misclassification	Assessment Statistics

Response	Event	Value	Training Frequency	Validation Frequency
Correct	1	True Positive	297	130
Incorrect	1	False Negative	35	18
Correct	0	True Negative	3,159	1,342
Incorrect	0	False Positive	9	10

Question 17: Insert a screenshot of your page.
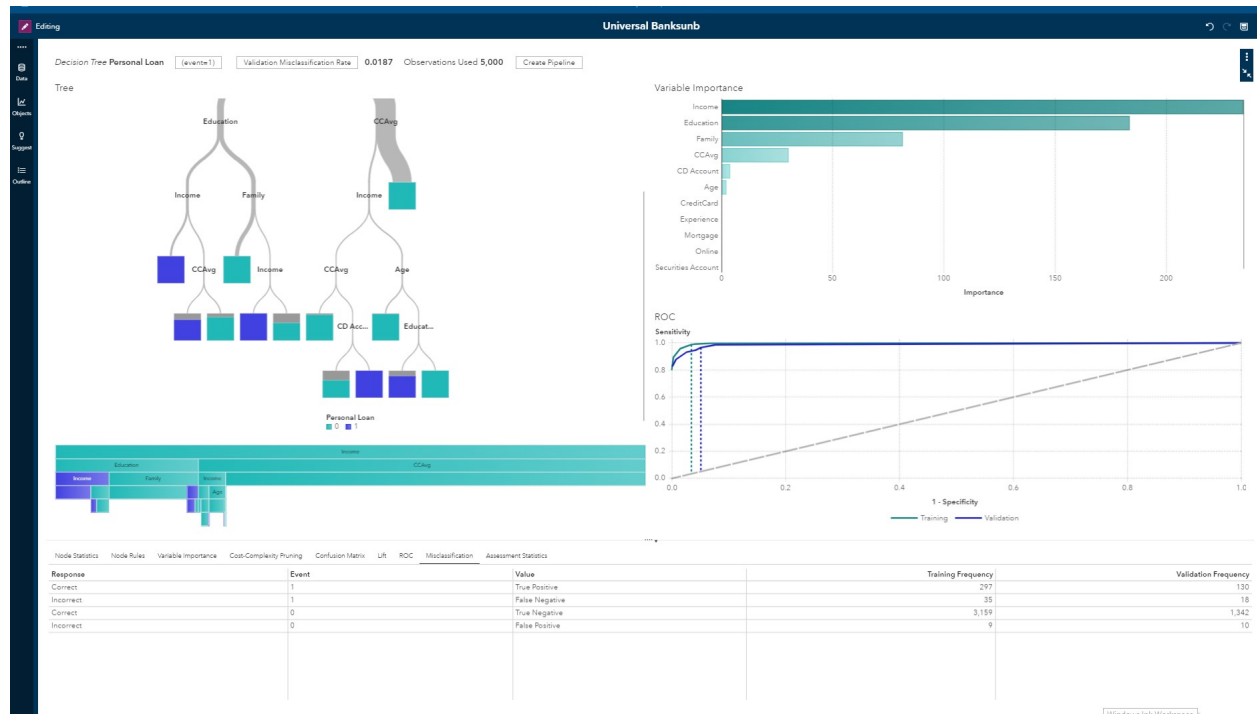
Question 18: What is the amount forecast for store 404 for 11/15/2016? What are the upper and lower confidence levels? What is the percent confidence level for your forecast?

The amount forecast for store 404 on 11/15/2016 is approximately 2348. The upper confidence level is 10845 and the lower confidence level is -6158. This forecast is using a 95% confidence level.

Question 19: Lower the confidence percent to 80%. What happens to the upper and lower bounds of the confidence levels?

The upper and lower bounds of the confidence level come closer to the amount forecast. The upper confidence level decreases to 7903 and the lower confidence level increases to -3215.

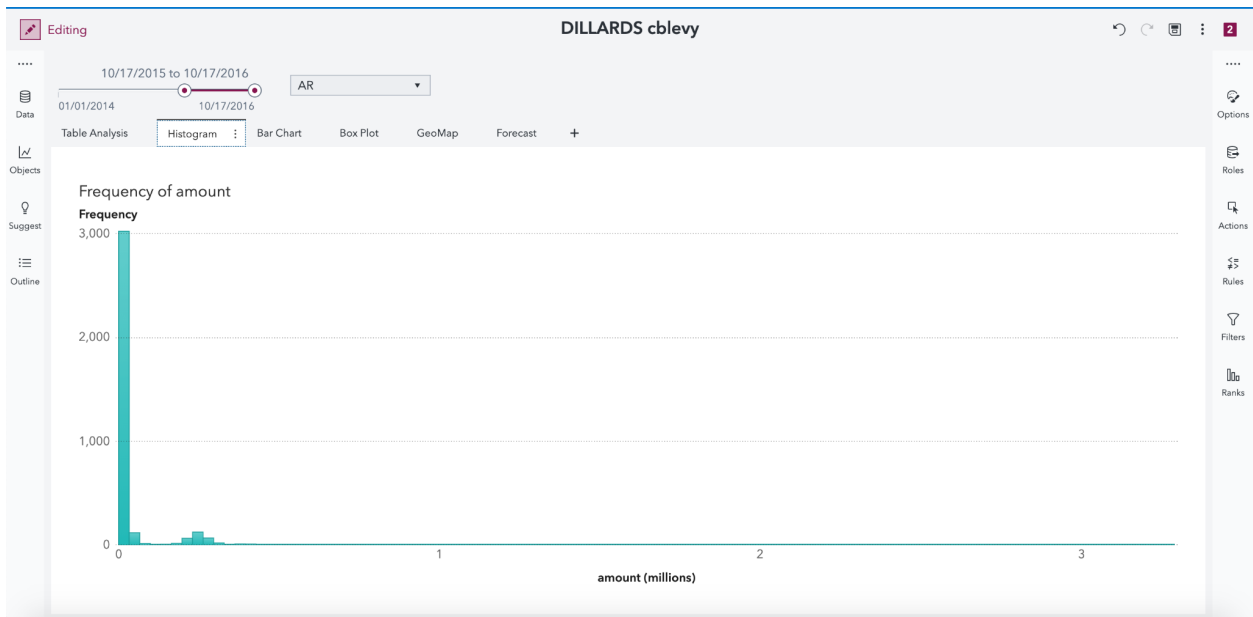## Question 20: Insert screenshots of your pages: Table Analysis, Histogram, Bar Chart, Box Plot, GeoMap, Forecast
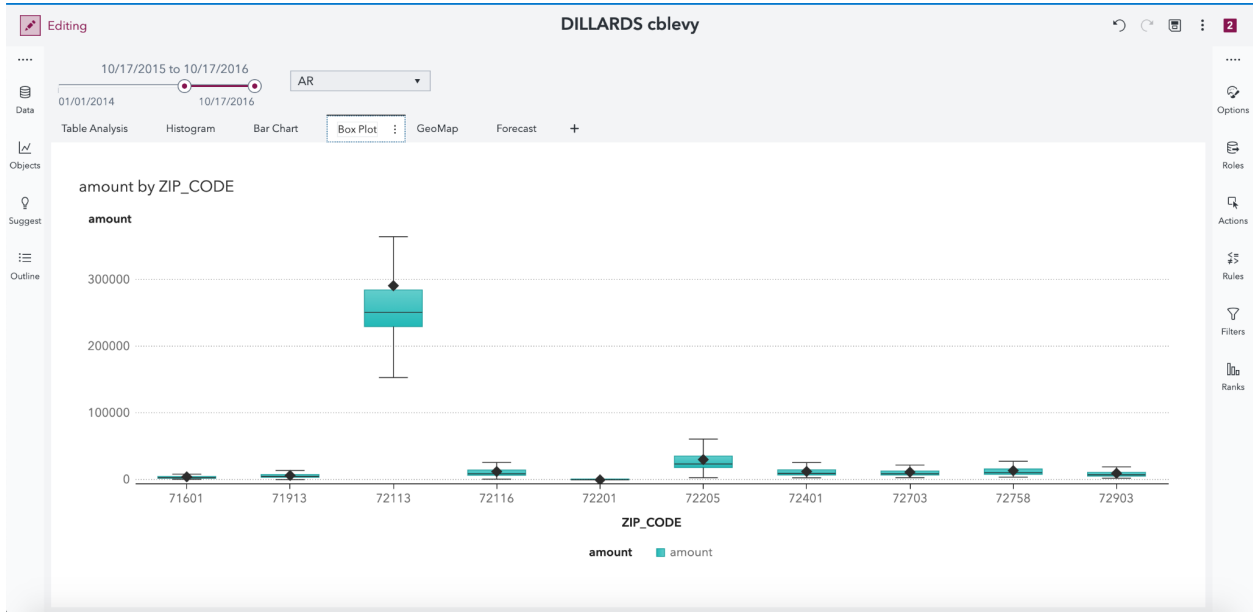
10/17/2015 to 10/17/2016

01/01/2014   10/17/2016
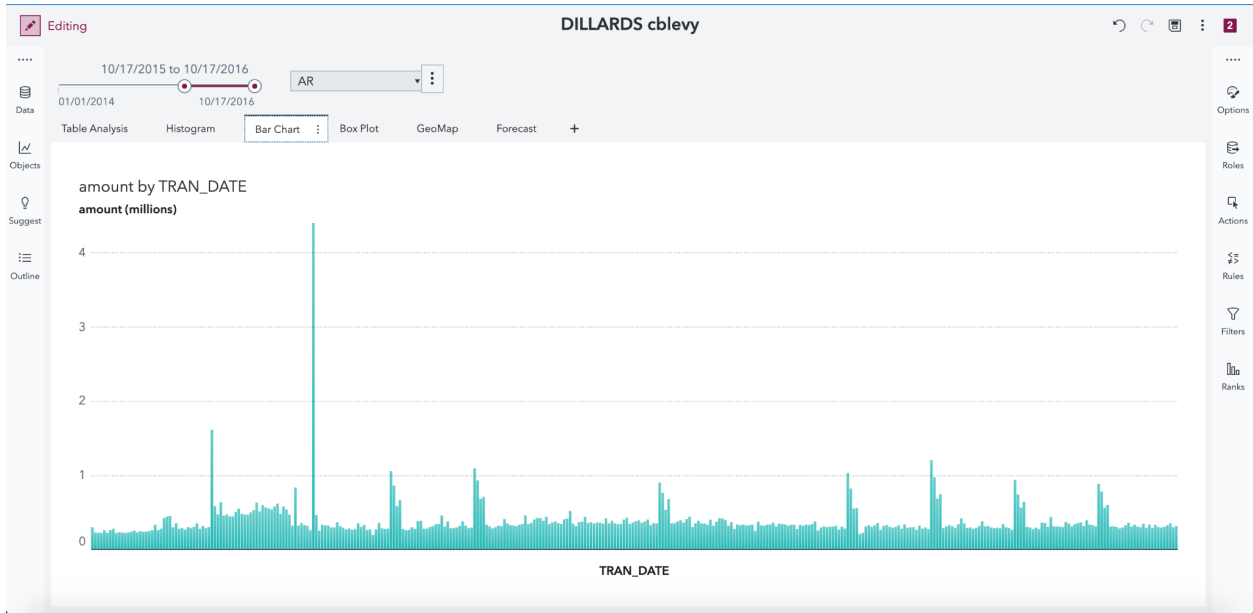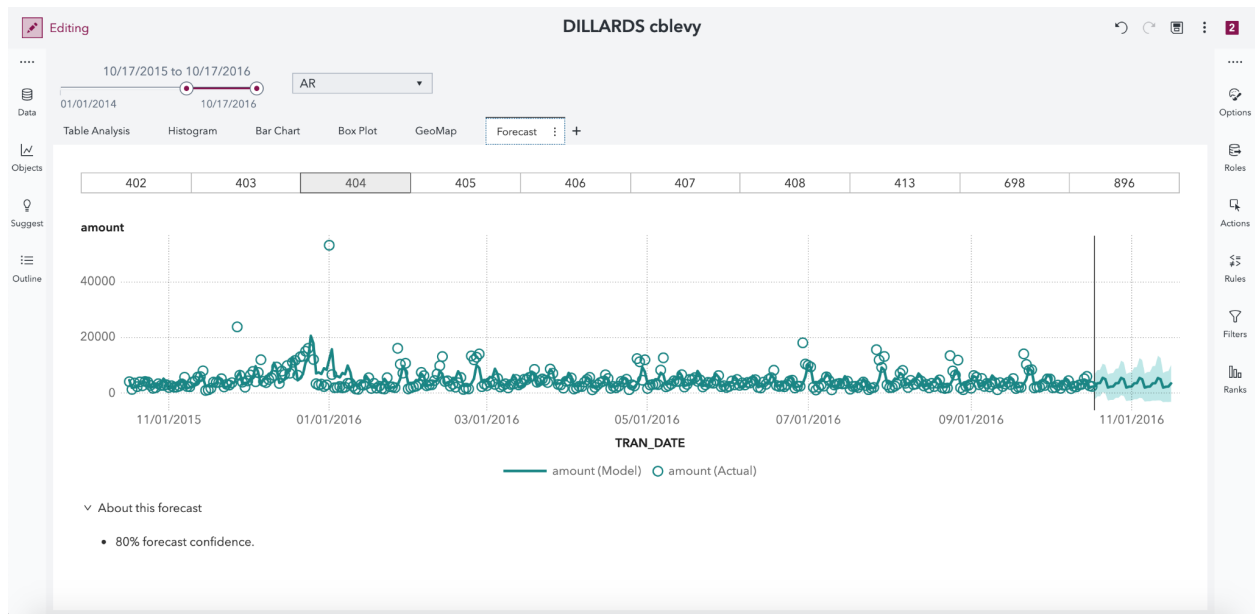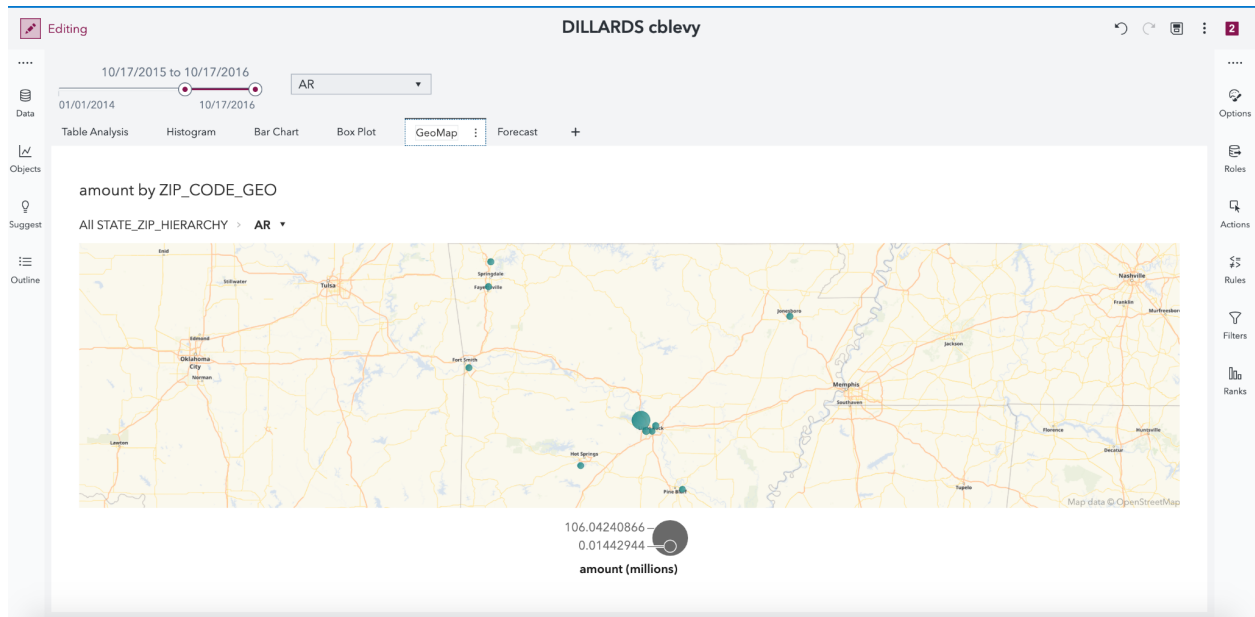
AR ▼

Table Analysis ⋮   Histogram   Bar Chart   Box Plot   GeoMap   Forecast   +

CITY	STATE ▲	TRAN_DATE	ZIP_CODE	amount	Frequency	STORE
FAYETTEVILLE	AR	04/25/2016	72703	6290.95	1	403
FAYETTEVILLE	AR	04/22/2016	72703	13161.48	1	403
FAYETTEVILLE	AR	09/24/2016	72703	18413.2	1	403
FAYETTEVILLE	AR	10/04/2016	72703	6474.39	1	403
FAYETTEVILLE	AR	05/25/2016	72703	8196.87	1	403
FAYETTEVILLE	AR	06/16/2016	72703	12900.71	1	403
FAYETTEVILLE	AR	02/25/2016	72703	22810.14	1	403
FAYETTEVILLE	AR	12/28/2015	72703	8348.24	1	403
FAYETTEVILLE	AR	12/19/2015	72703	35217.16	1	403
FAYETTEVILLE	AR	12/01/2015	72703	10187.33	1	403
FAYETTEVILLE	AR	05/13/2016	72703	11814.24	1	403
FAYETTEVILLE	AR	10/19/2015	72703	7117.78	1	403
FAYETTEVILLE	AR	10/31/2015	72703	12623.96	1	403
FAYETTEVILLE	AR	01/23/2016	72703	15911.49	1	403
FAYETTEVILLE	AR	08/04/2016	72703	8115.87	1	403
FAYETTEVILLE	AR	02/26/2016	72703	28631.8	1	403
FAYETTEVILLE	AR	09/07/2016	72703	6096.67	1	403
FAYETTEVILLE	AR	02/27/2016	72703	36127.6	1	403
FAYETTEVILLE	AR	11/01/2015	72703	6254.77	1	403

10/17/2015 to 10/17/2016

01/01/2014   10/17/2016

AR ▼

Table Analysis   Histogram ⋮   Bar Chart   Box Plot   GeoMap   Forecast   +

Frequency of amount

Editing — DILLARDS cblevy

10/17/2015 to 10/17/2016
01/01/2014 — 10/17/2016
AR

Table Analysis | Histogram | Bar Chart | Box Plot | GeoMap | Forecast | +

amount by TRAN_DATE
amount (millions)

TRAN_DATE

Editing — DILLARDS cblevy

10/17/2015 to 10/17/2016
01/01/2014 — 10/17/2016
AR

Table Analysis | Histogram | Bar Chart | Box Plot | GeoMap | Forecast | +

amount by ZIP_CODE
amount

300000

200000

100000

0

71601 71913 72113 72116 72201 72205 72401 72703 72758 72903
ZIP_CODE

amount ▪ amount

Save it as *group#_final_project.* **Submit on Blackboard.**