ANZ TASK 1 - Exploratory Data Analysis

OBSERVATIONS (NEED FURTHER INVESTIGATION)

Observation 1 : Card Flag Status

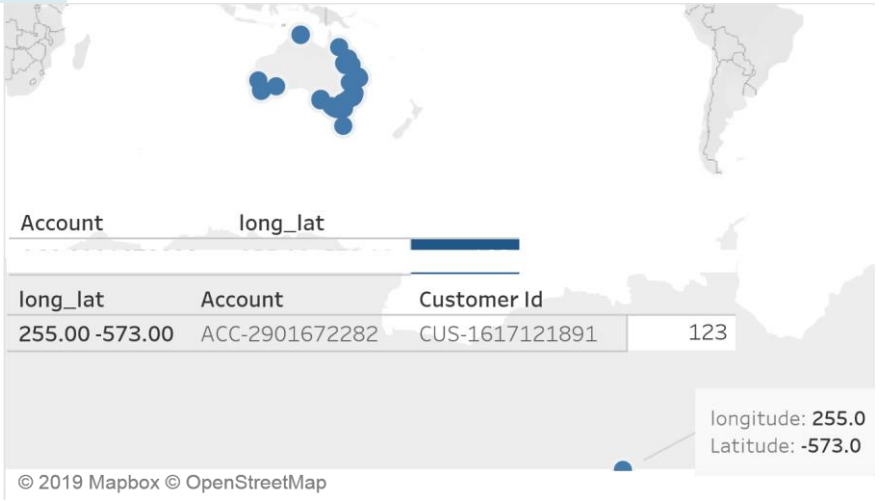
Movement	Txn Descripti..	Count of Card Pres..	Card Present Flag	
Grand Total				12,043
credit	PAY/SALARY	0	Null	883
debit	INTER BANK	0	Null	742
	PAYMENT	0	Null	2,600
	PHONE BA..	0	Null	101
	POS	758	0	758
		3,025	1	3,025
	SALES-POS	765	0	765
		3,169	1	3,169

Comment:

There are some POS and SALES-POS transactions with credit card flag = 0.

Both cases are closed to 25% (POS: 758/3025; SALES-POS : 765/3169). Could it be due to some operation fault?

Observation 3 : Customer Location



Comment :

Customer Daniel (Acc. No 2901672282) have invalid Geocode for his location.

Observation 2: Min Transaction Amt

Movement	Txn Description	
credit	PAY/SALARY	576.0
debit	INTER BANK	16.0
	PAYMENT	15.0
	PHONE BANK	21.0
	POS	0.1
	SALES-POS	0.1

Comment:

There are 5 transaction values with very little amount of 0.1 .

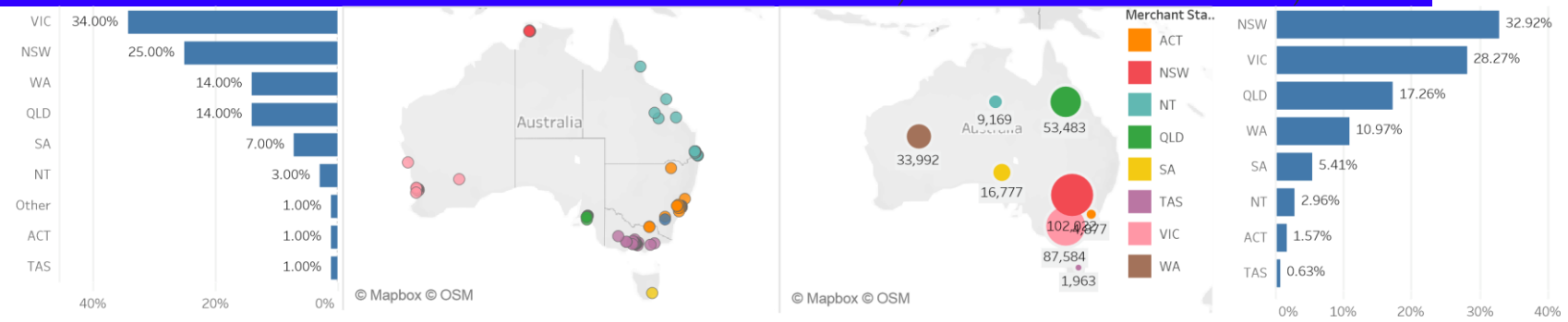
TxnDescription	AmountMin.	Count
POS	0.100	2.000
SALES-POS	0.100	3.000

It seems very strange to have such low value transaction. Probably need to investigate

Probably need to investigate why is that so?

Customer Spending & Merchant Analysis (By State)

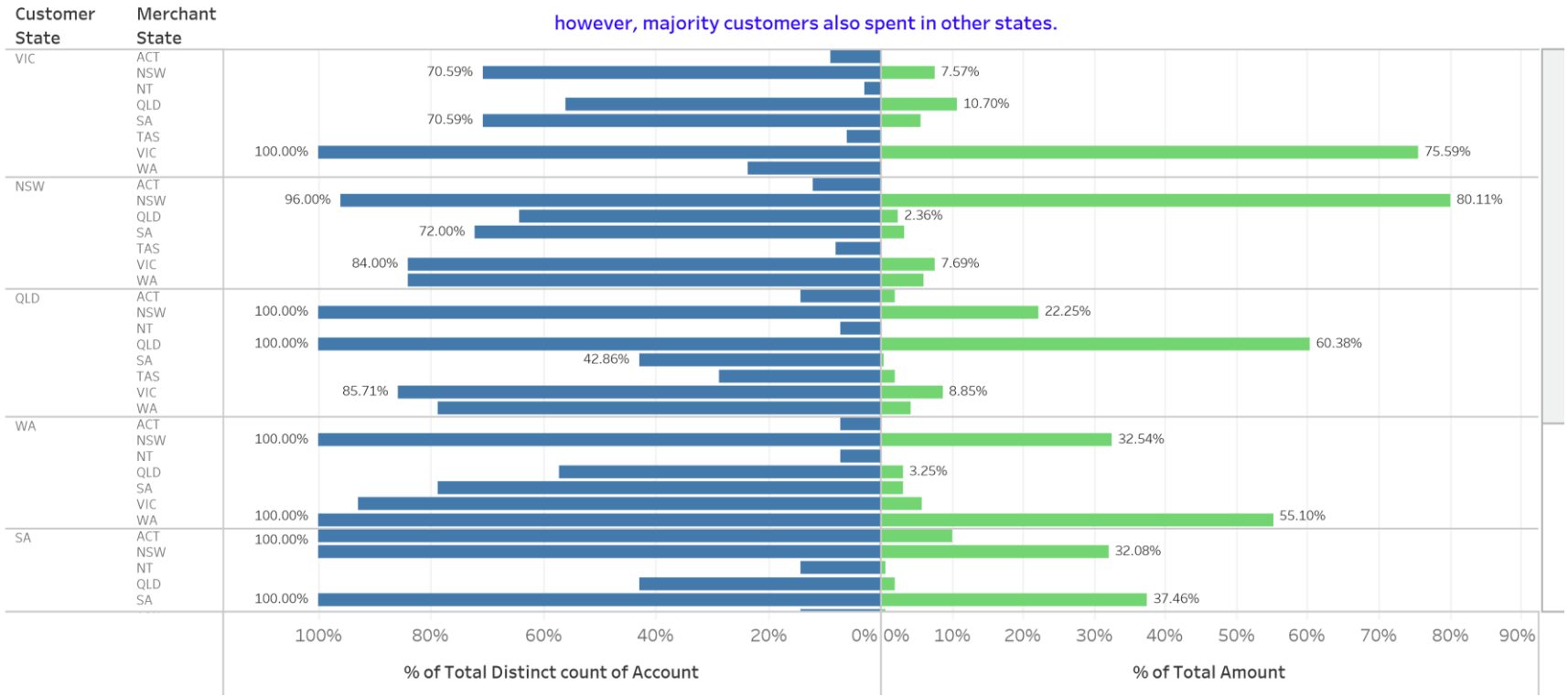
Customer Distribution By State | Customer State | MERCHANT DIST (SALE POS & POS) | MERCHANT (SALE POS & POS)



Largest group of ANZ's Account Holders are from VIC (34%), followed by NSW (25%). However, Most spending to merchant in NSW (32.92%), which is mainly contributed by local customers and portion from customers of other states . See Bar Chart Below.

Customer Spending Distribution By State (SALE-POS & POS)

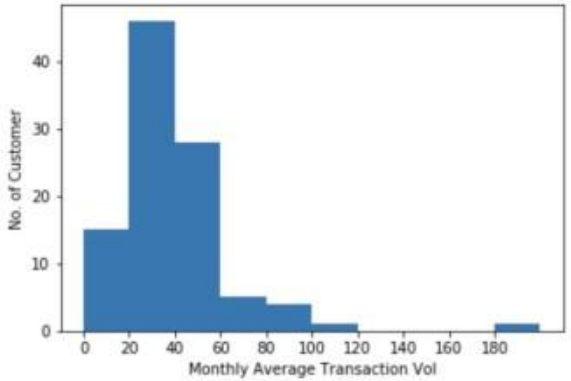
Most customers spent in their own state (except ACT and TAS),



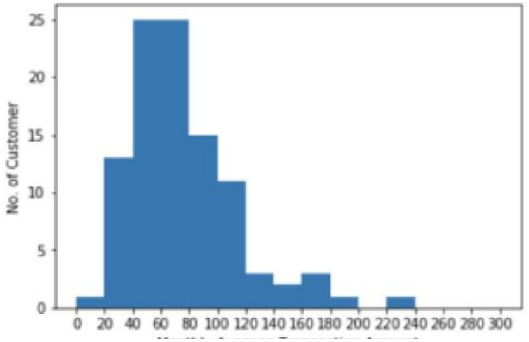
Monthly Average Transaction Volume and Amount (by customers)

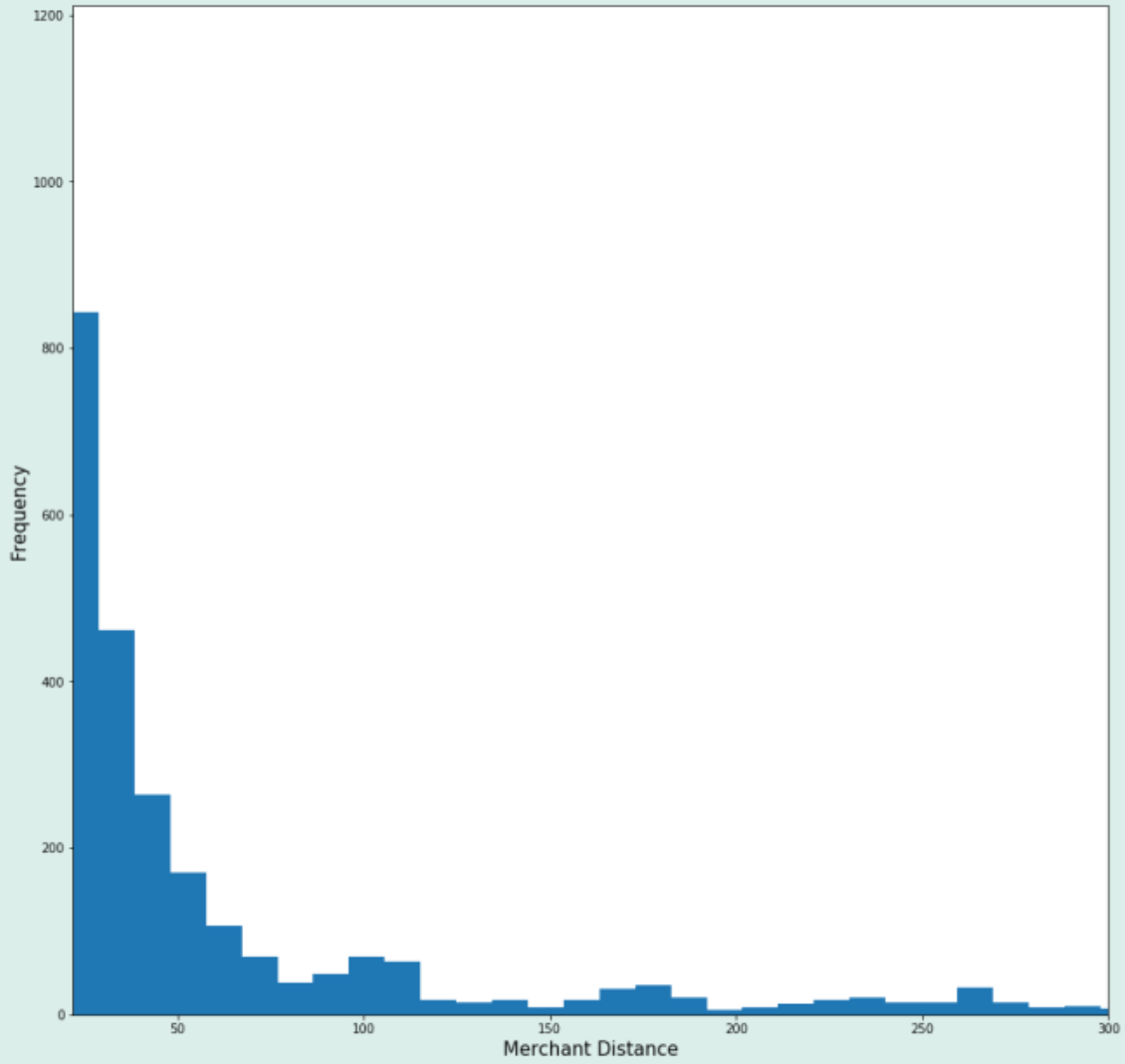
Please find Jupyter notebook link below for the codes and original image file

In [247]:
```
# Calculate the Monthly Average Transaction Vol

Vol = df.amount.groupby(df.account)
Ave_Vol = Vol.count()/3

plt.hist(Ave_Vol, bins = range(0,220,20))
plt.xlabel('Monthly Average Transaction Vol')
plt.ylabel('No. of Customer')
plt.xticks(np.arange(0, max(Ave_Vol)+1, 20.0))
plt.show()
```



In [248]:
```
# Calculate the Monthly Average Transactiojn Amount

Ave_Amt = Vol.mean()/3

plt.hist(Ave_Amt,bins = range(0,320,20))
plt.xlabel('Monthly Average Transaction Amount')
plt.ylabel('No. of Customer')
plt.xticks(np.arange(0, 320, 20.0))
plt.show()
```





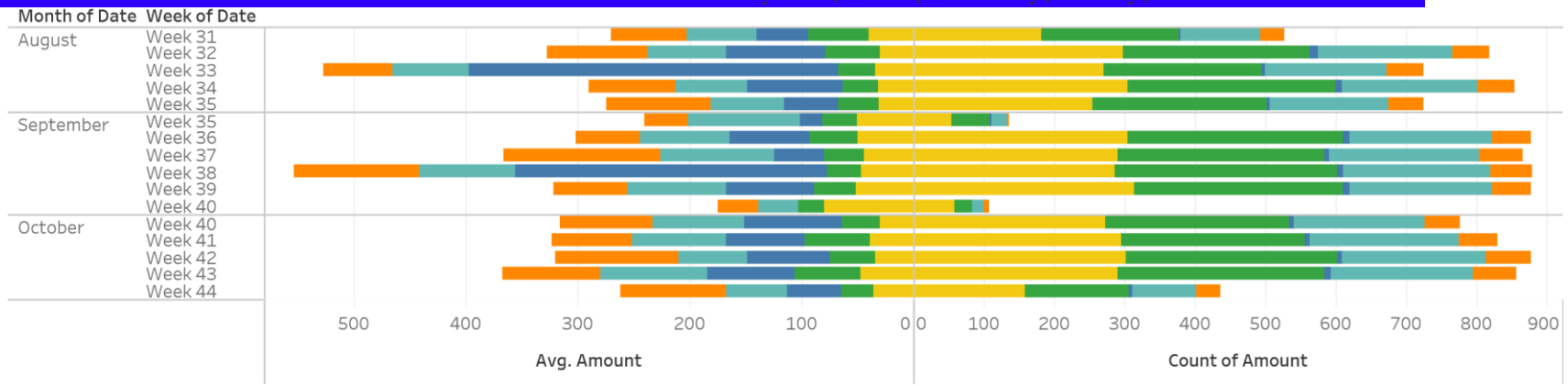
Customer Distance to Merchant Distribution

Please find Jupyter notebook link below for the codes and original image file

https://github.com/YANGTN/ANZ_TASK1/blob/master/ANZ_TASK%201.ipynb

DATA SEGMENTATION (By Week & By HR)

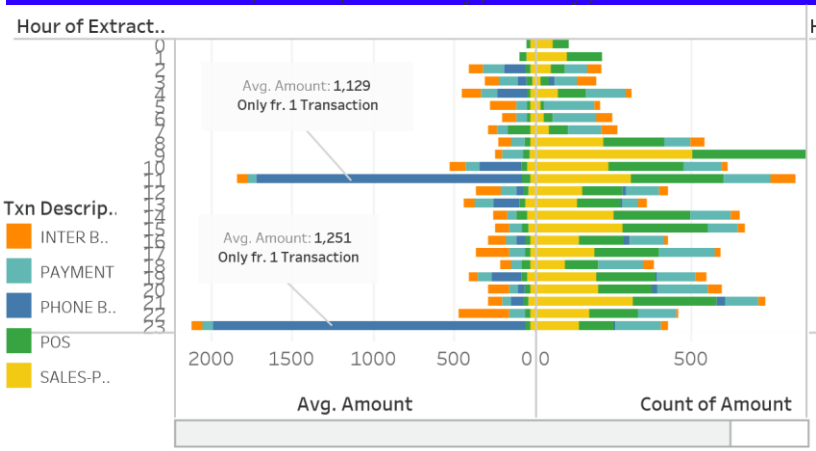
AVG DEBIT TRANSACTION_Amt & Volume by Acc/ Wk (Omit Pay/Salary)



Comments:

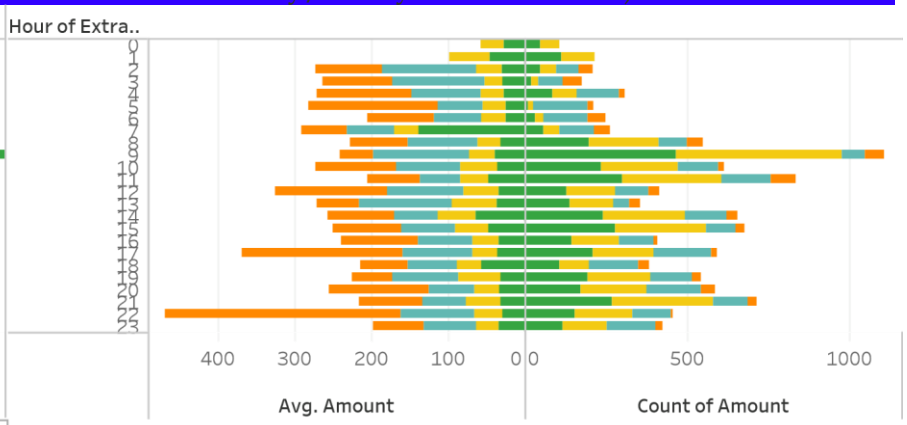
Most debit transactions seem to peak during middle of the month and low at end of the month

AVG DEBIT TRANSACTION_Amt & Volume by Acc/ HR (Omit Pay/Salary)



Comments:

AVG DEBIT TRANSACTION_Amt & Vol by Acc/ HR (Omit Pay/Salary & Phone Bank)

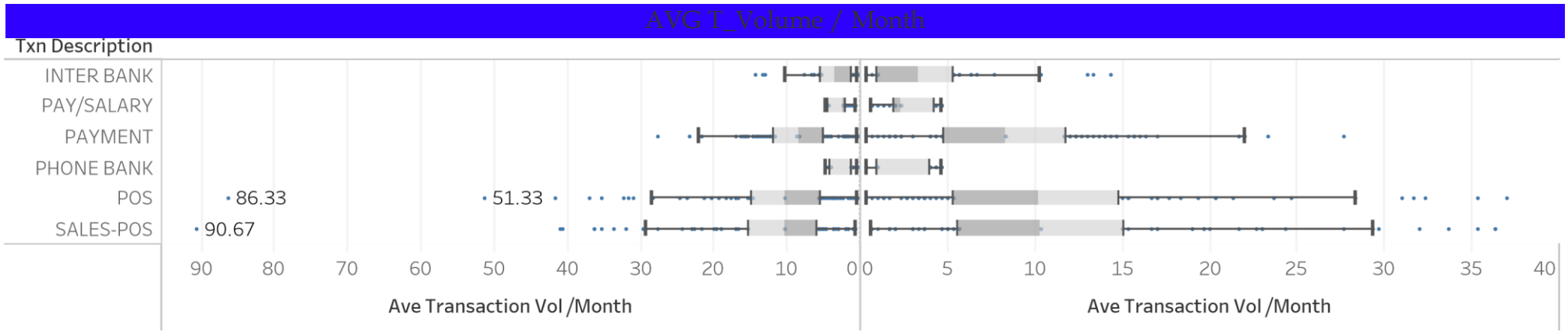


Comments:

Phone Bank Average for 11hr and 23 hr are only single transaction, which is It is observed that average transaction volume surged at certain timing: not good representation of the samples.

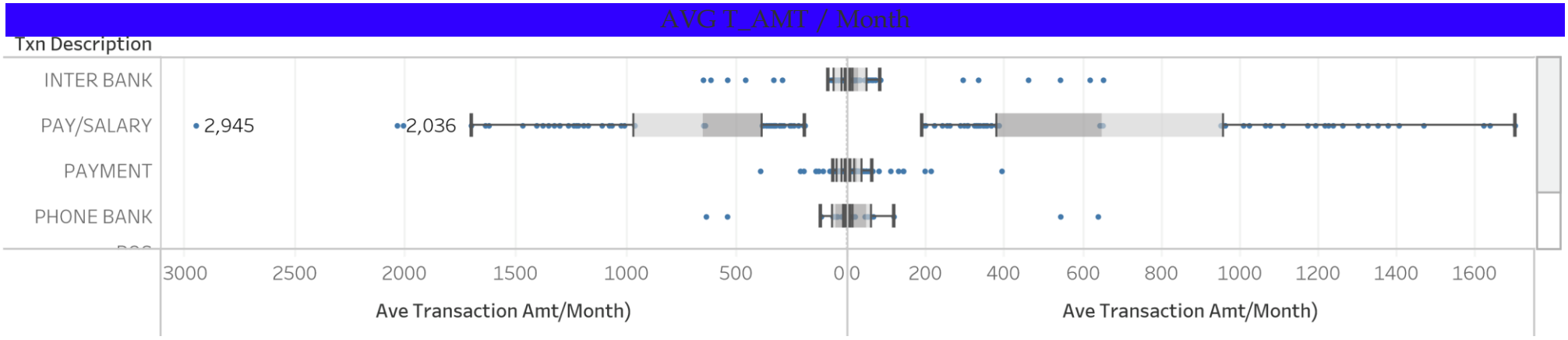
Thus, decided to remove Phone bank 9 AM : coincide with start of day; for analysis (refer chart on right) 11AM, 2PM : Lunch and tea break 9PM : after work

TRANSACTION ANALYSIS



Comments:

Box Plot with outliers taken out from the observation range enabled easier visual analysis of the statistics description. SALES-POS and POS are the major mode of debit transaction.

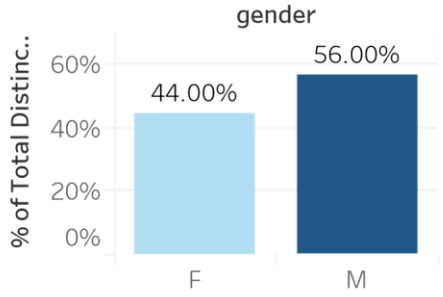


Comments:

Box Plot with outliers taken out from the observation range enabled easier visual analysis of the statistics description. Most highest Amount of transaction is due to Pay/Salary. In order to analyze debit transaction, transaction due to Pay/ Salary should be omited.

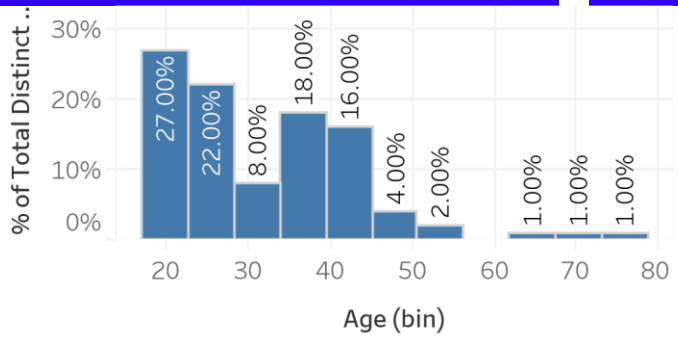
Customer Demographic (Extra Materials for Reference)

Gender Dist



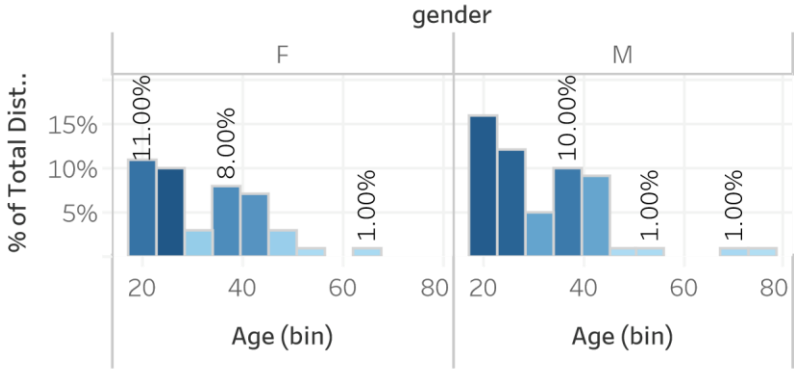
Comments:
There are slightly more male than fe..

Age Dist (All)

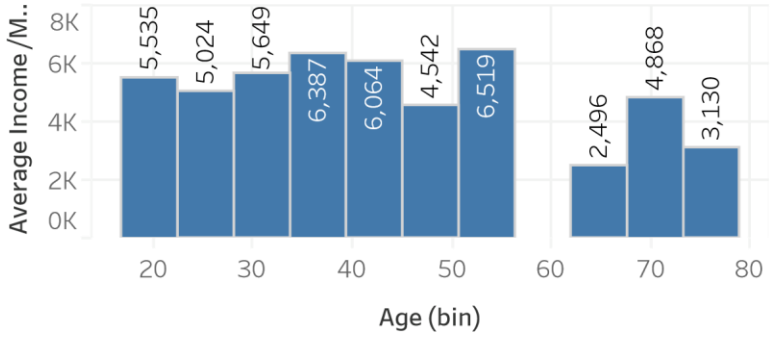


Comments:
Majority of customers are in the age of 45 and below...

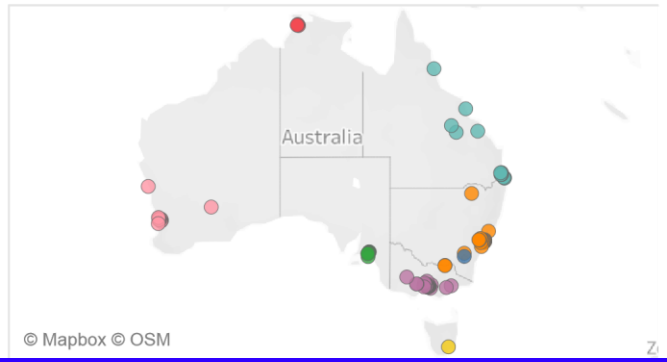
Age Dist (By Gender)



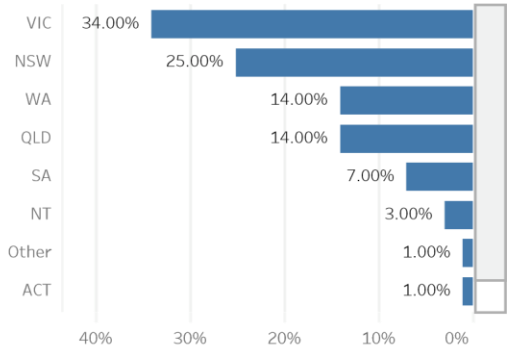
AVG Income



Customer State



Customer Distribution By State

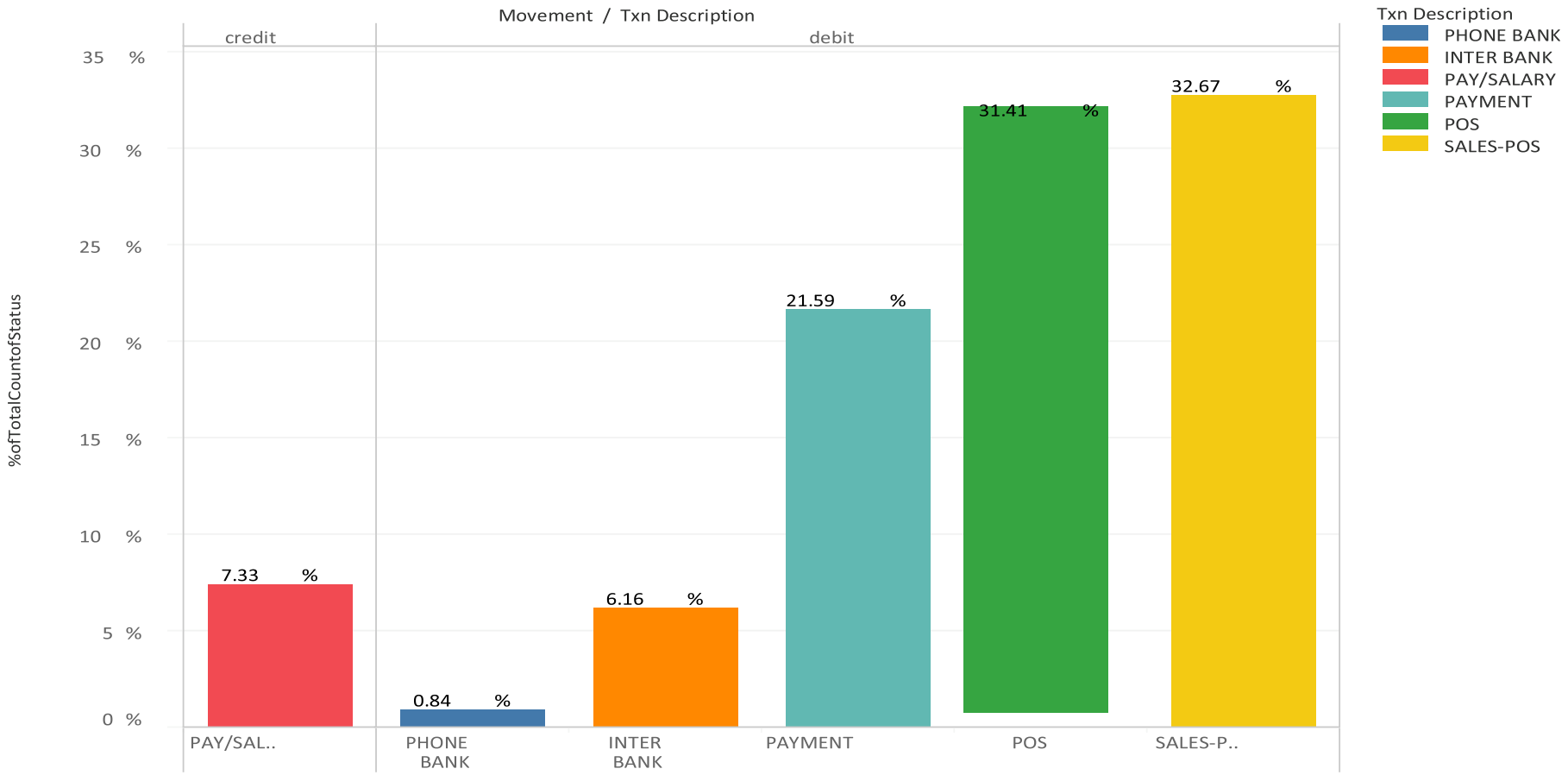


Comments:

Bar charts above shows that no significant correlation between age and average monthly income for ANZ custom..

Comments:

Most customers are from Victoria (34), followed by NSW (25), QLD & WA (14 respectively).



Transactions by type description broken down by Movement. Color shows details about Txn Description.