ANZ TASK 1 - Exploratory Data Analysis

File created on: 6/8/2019 9:14:57 AM

Prepared by : Teik Ning Yang (RTC)

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 2: Min Transaction Amt

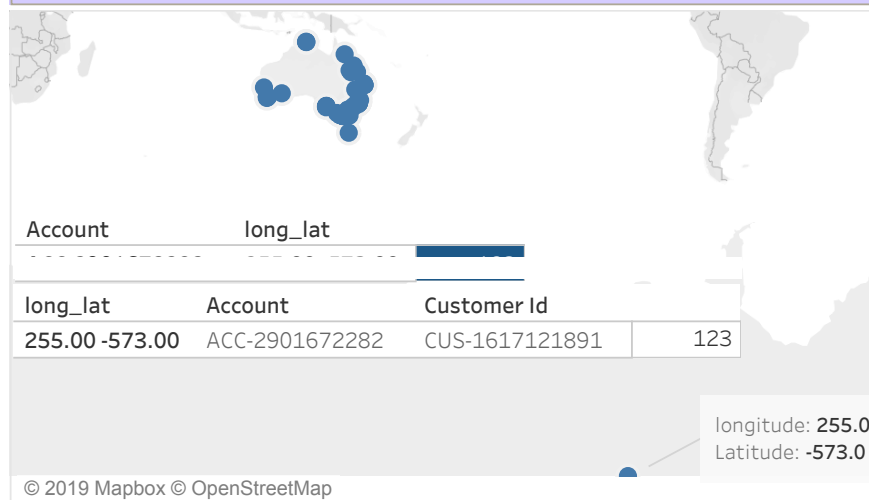| Movement	Txn Description	
credit	PAY/SALARY	576.0
debit	INTER BANK	16.0
	PAYMENT	15.0
	PHONE BANK	21.0
	POS	0.1
	SALES-POS	0.1

Comment:
There are 5 transaction values with very little amount of 0.1 .
It seems very strange to have such low value transaction.
Probably need to investigate

Txn Description	Min. Amount	Count
POS	0.100	2.000
SALES-POS	0.100	3.000

Observation 3 : Customer Location

Account	long_lat	

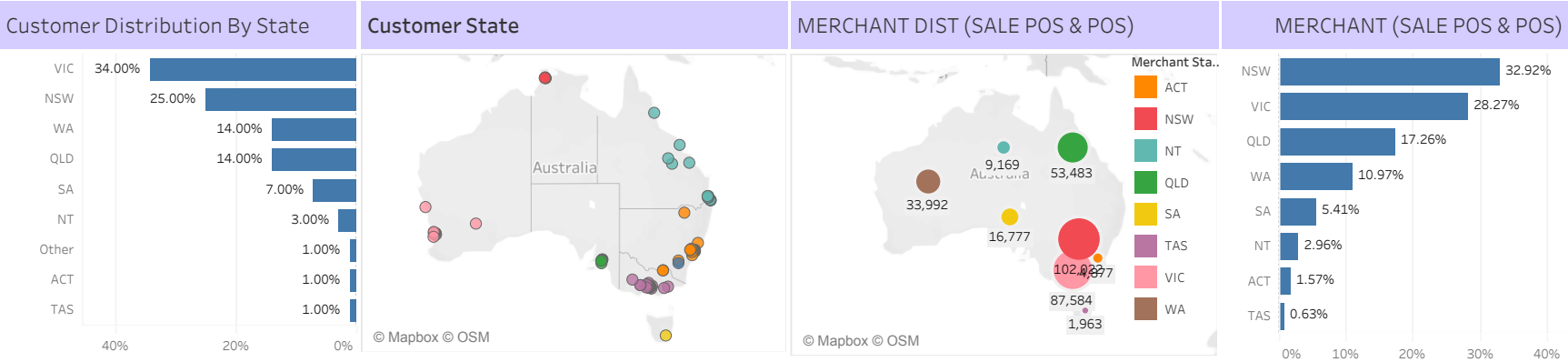| long_lat	Account	Customer Id	
255.00 -573.00	ACC-2901672282	CUS-1617121891	123

longitude: **255.0**
Latitude: **-573.0**

© 2019 Mapbox © OpenStreetMap

Comment :
Customer Daniel (Acc. No 2901672282) have invalid Geocode for his location.
Probably need to investigate why is that so?

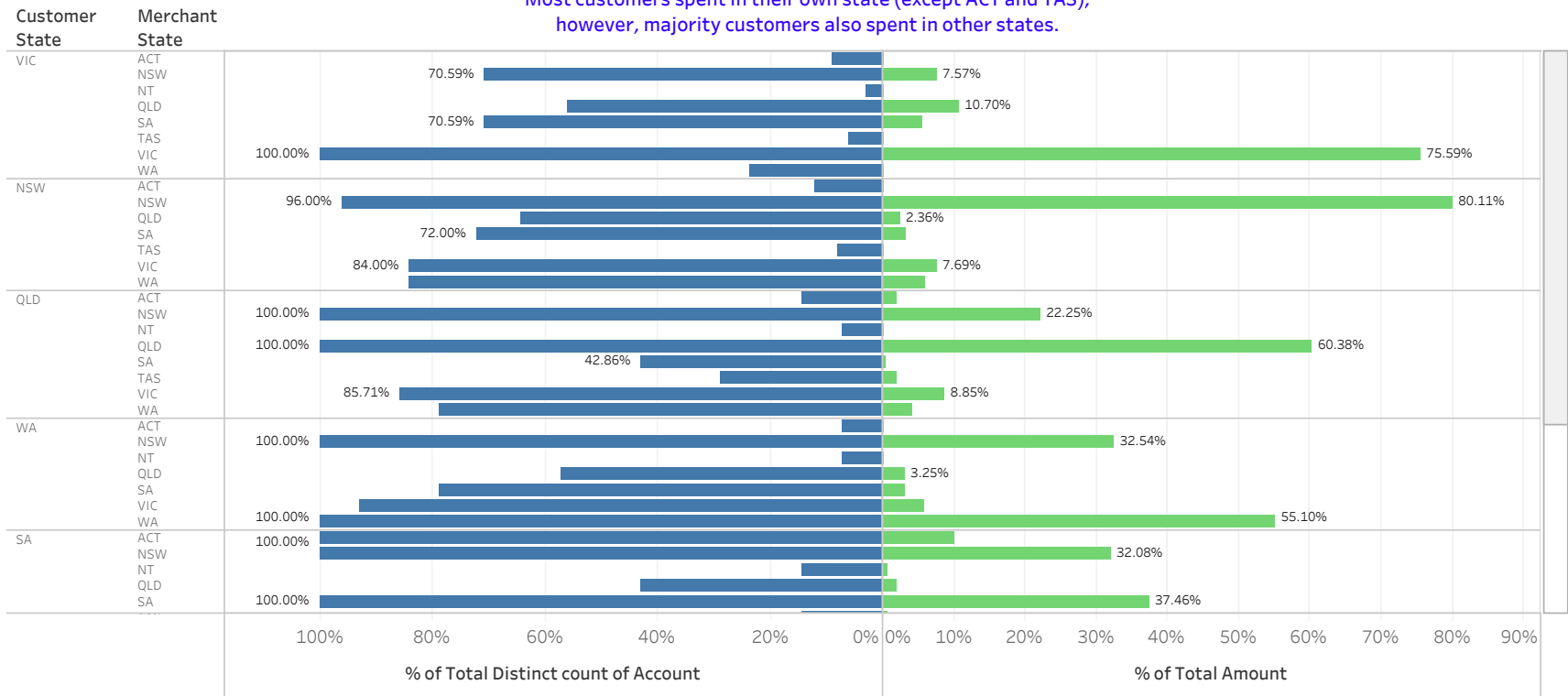# Customer Spending & Merchant Analysis (By State)

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

Customer Distribution By State

- VIC — 34.00%
- NSW — 25.00%
- WA — 14.00%
- QLD — 14.00%
- SA — 7.00%
- NT — 3.00%
- Other — 1.00%
- ACT — 1.00%
- TAS — 1.00%

(axis: 40% 20% 0%)

Customer State

Australia
© Mapbox © OSM

MERCHANT DIST (SALE POS & POS)

Merchant Sta..
- ACT
- NSW
- NT
- QLD
- SA
- TAS
- VIC
- WA

9,169
53,483
33,992
16,777
102,022
9,877
87,584
1,963

Australia
© Mapbox © OSM

MERCHANT (SALE POS & POS)

- NSW — 32.92%
- VIC — 28.27%
- QLD — 17.26%
- WA — 10.97%
- SA — 5.41%
- NT — 2.96%
- ACT — 1.57%
- TAS — 0.63%

(axis: 0% 10% 20% 30% 40%)

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),
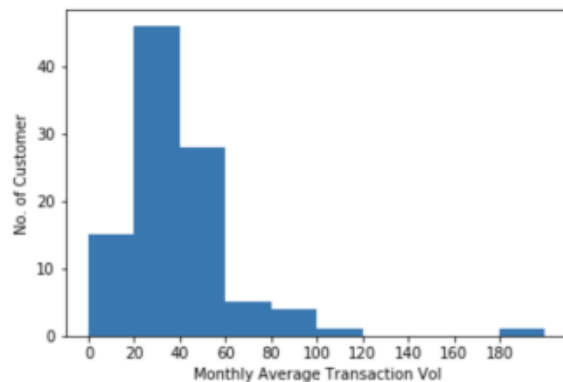however, majority customers also spent in other states.

Customer State	Merchant State	% of Total Distinct count of Account	% of Total Amount
VIC	ACT		
	NSW	70.59%	7.57%
	NT		
	QLD		10.70%
	SA	70.59%	
	TAS		
	VIC	100.00%	75.59%
	WA		
NSW	ACT		
	NSW	96.00%	80.11%
	QLD		2.36%
	SA	72.00%	
	TAS		
	VIC	84.00%	7.69%
	WA		
QLD	ACT		
	NSW	100.00%	22.25%
	NT		
	QLD	100.00%	60.38%
	SA	42.86%	
	TAS		
	VIC	85.71%	8.85%
	WA		
WA	ACT		
	NSW	100.00%	32.54%
	NT		
	QLD		3.25%
	SA		
	VIC		
	WA	100.00%	55.10%
SA	ACT	100.00%	
	NSW		32.08%
	NT		
	QLD		
	SA	100.00%	37.46%

(X-axis left: 100% 80% 60% 40% 20% 0% — % of Total Distinct count of Account)
(X-axis right: 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% — % of Total Amount)

Monthly Average Transaction Volume and Amount (by customers)

In [247]:
```python
# 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()
```
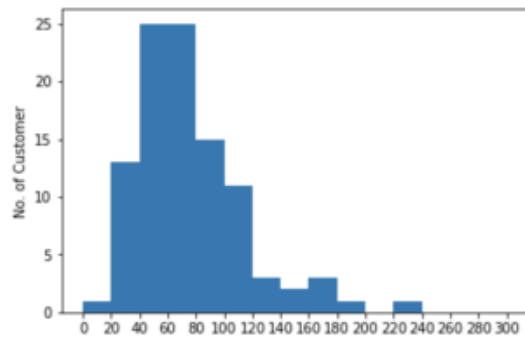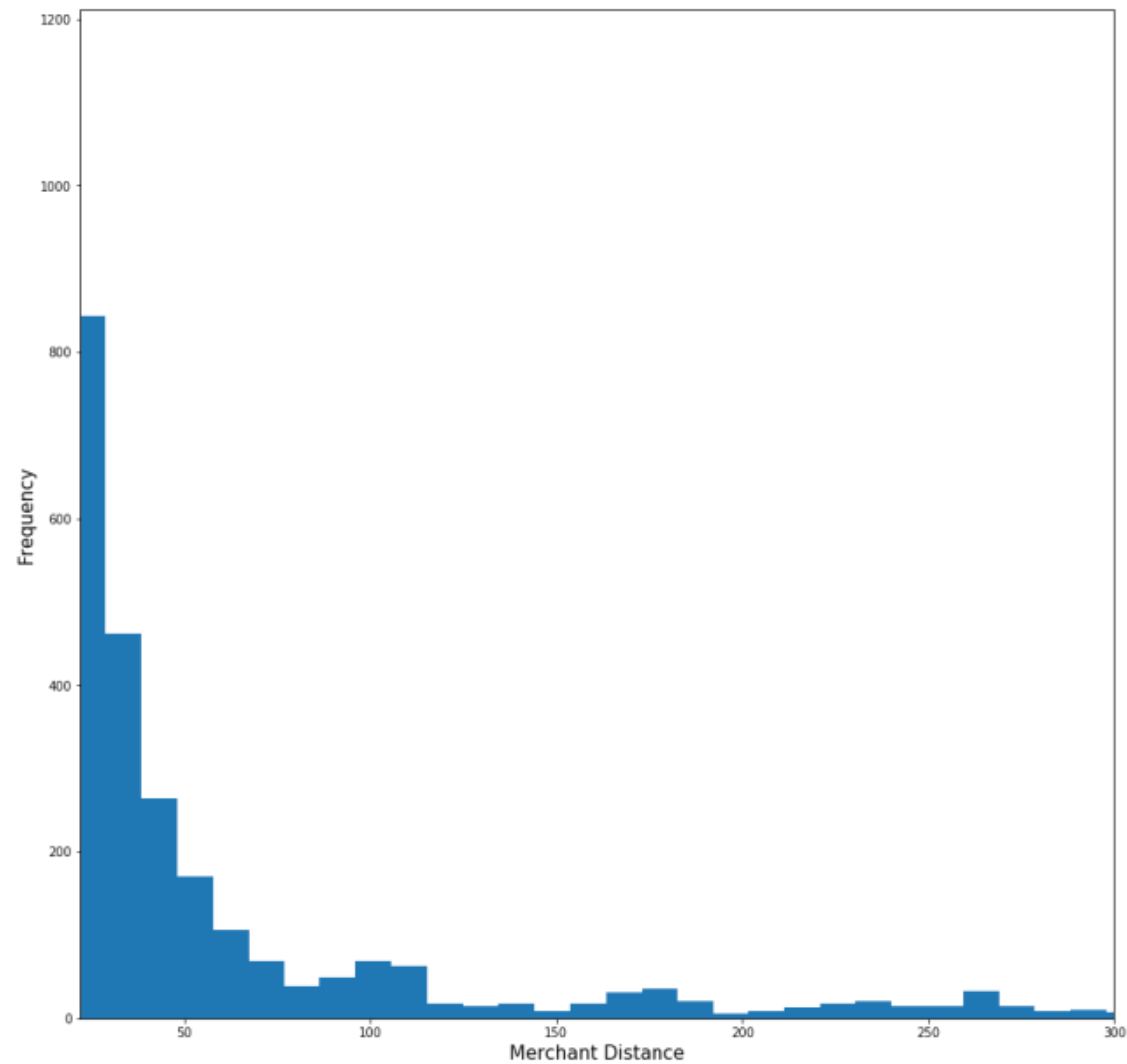
Please find Jupyter notebook link below for the codes and original image file

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

In [248]:
```python
# 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()
```

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

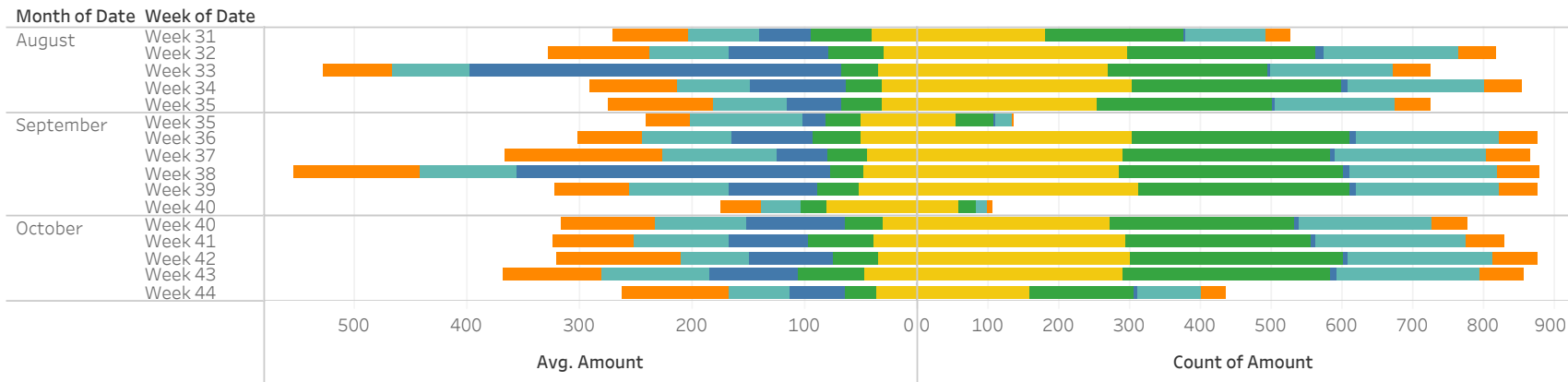# Customer Distance to Merchant Distribution

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)

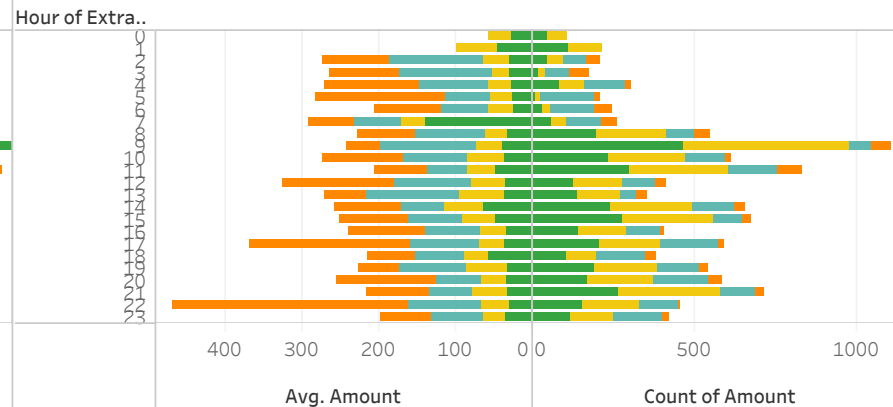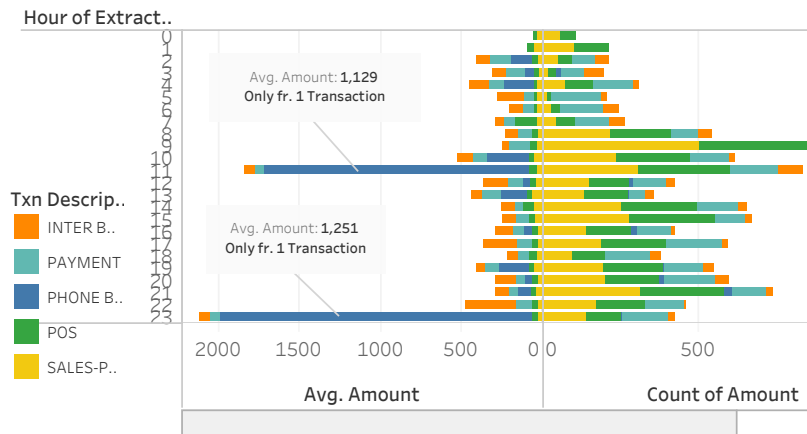## AVG DEBIT TRANSACTION_Amt & Volume by Acc/ Wk (Omit Pay/Salary)

Month of Date	Week of Date
August	Week 31
	Week 32
	Week 33
	Week 34
	Week 35
September	Week 35
	Week 36
	Week 37
	Week 38
	Week 39
	Week 40
October	Week 40
	Week 41
	Week 42
	Week 43
	Week 44

Avg. Amount: 500 400 300 200 100 0
Count of Amount: 0 100 200 300 400 500 600 700 800 900

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)

Hour of Extract..

Avg. Amount: 1,129
Only fr. 1 Transaction

Avg. Amount: 1,251
Only fr. 1 Transaction

Txn Descrip..
- INTER B..
- PAYMENT
- PHONE B..
- POS
- SALES-P..

Avg. Amount: 2000 1500 1000 500 0
Count of Amount: 0 500

Comments:

The Phone Bank Average for 11hr and 23 hr are only single transaction, which is not good representation of the samples. Thus, decided to remove Phone bank for analysis (refer chart on right)
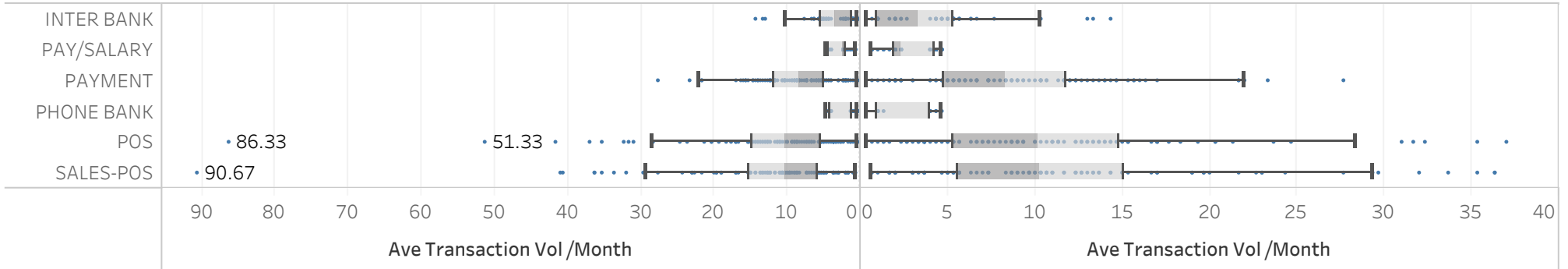
AVG DEBIT TRANSACTION_Amt & Vol by Acc/ HR (Omit Pay/Salary & Phone Bank)

Hour of Extra..

Avg. Amount: 400 300 200 100 0
Count of Amount: 0 500 1000

Comments:

It is observed that average transaction volume surged at certain timing:

9 AM : coincide with start of day;

11AM, 2PM : Lunch and tea break

9PM : after work

TRANSACTION ANALYSIS

AVG T_Volume / Month

Txn Description

INTER BANK
PAY/SALARY
PAYMENT
PHONE BANK
POS   • 86.33   • 51.33
SALES-POS   • 90.67

90 80 70 60 50 40 30 20 10 0   0 5 10 15 20 25 30 35 40

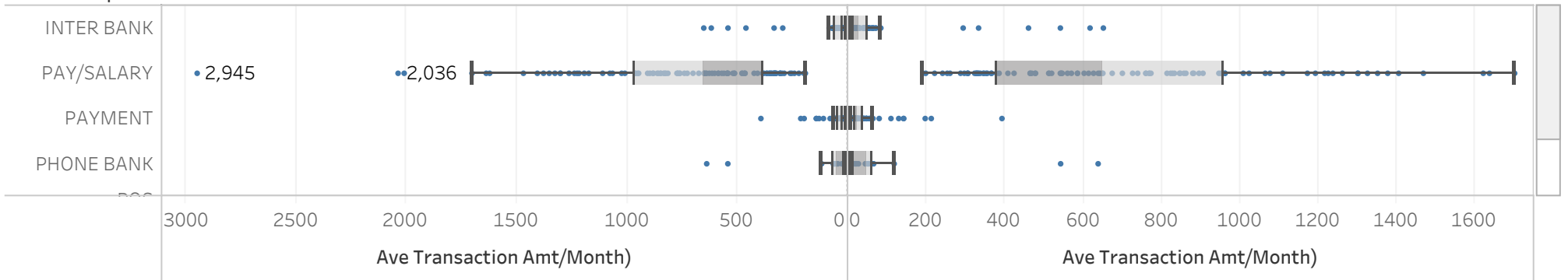Ave Transaction Vol /Month   Ave Transaction Vol /Month

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.

AVG T_AMT / Month

Txn Description

INTER BANK
PAY/SALARY   • 2,945   • 2,036
PAYMENT
PHONE BANK

3000 2500 2000 1500 1000 500 0 0 200 400 600 800 1000 1200 1400 1600

Ave Transaction Amt/Month)   Ave Transaction Amt/Month)

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.
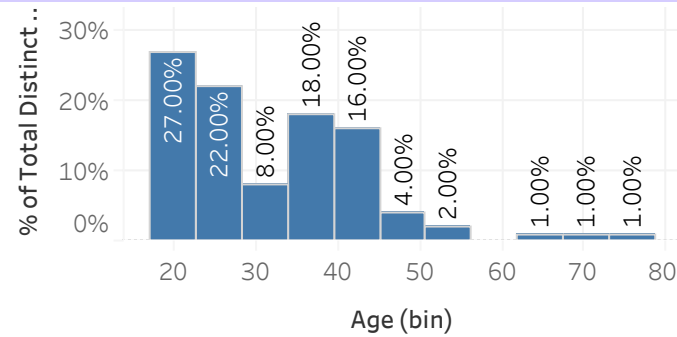
Customer Demographic (Extra Materials for Reference)
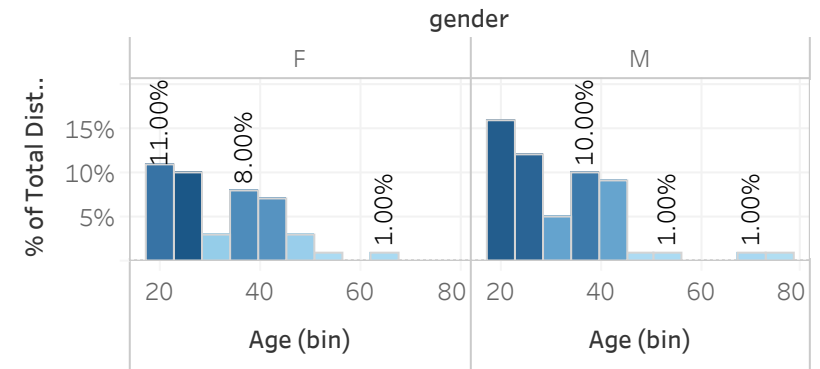
Gender Dist

gender

44.00%   56.00%

% of Total Distinc..

F   M
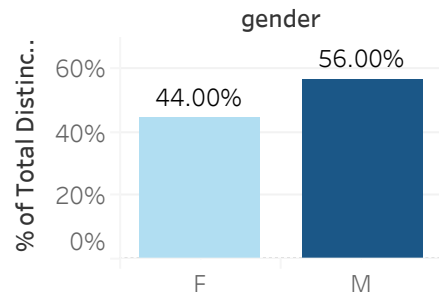
Comments:
There are slightly more male than fe..

Age Dist (All)

% of Total Distinct..

27.00% 22.00% 8.00% 18.00% 16.00% 4.00% 2.00% 1.00% 1.00% 1.00%

Age (bin)

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

Age Dist (By Gender)

gender

F   M

% of Total Dist..

11.00% 8.00% 1.00%   10.00% 1.00% 1.00%

Age (bin)   Age (bin)

AVG Income

Average Income / M..

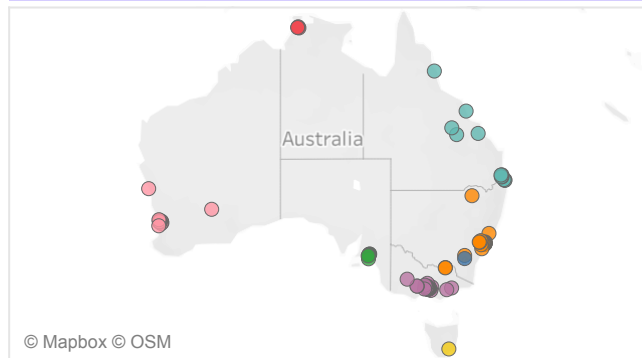5,535 5,024 5,649 6,387 6,064 4,542 6,519 2,496 4,868 3,130

Age (bin)
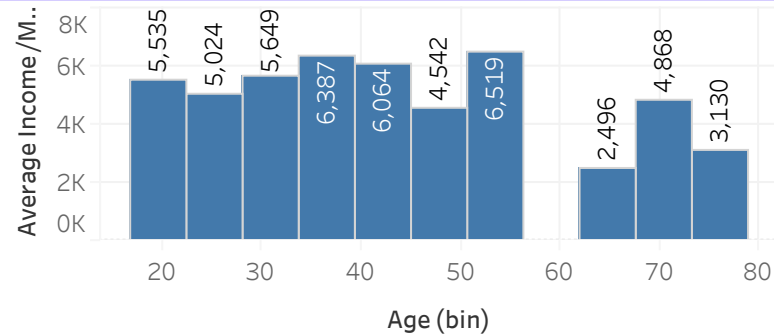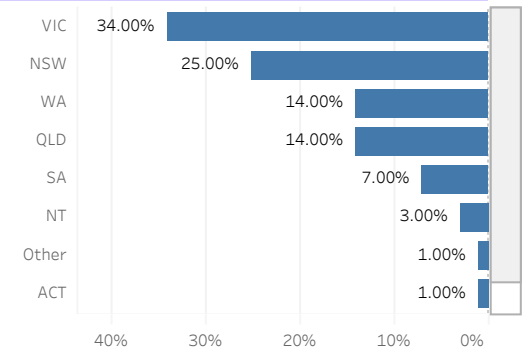
Comments:
Bar charts above shows that no significant correlation between age and average monthly income for ANZ custom..
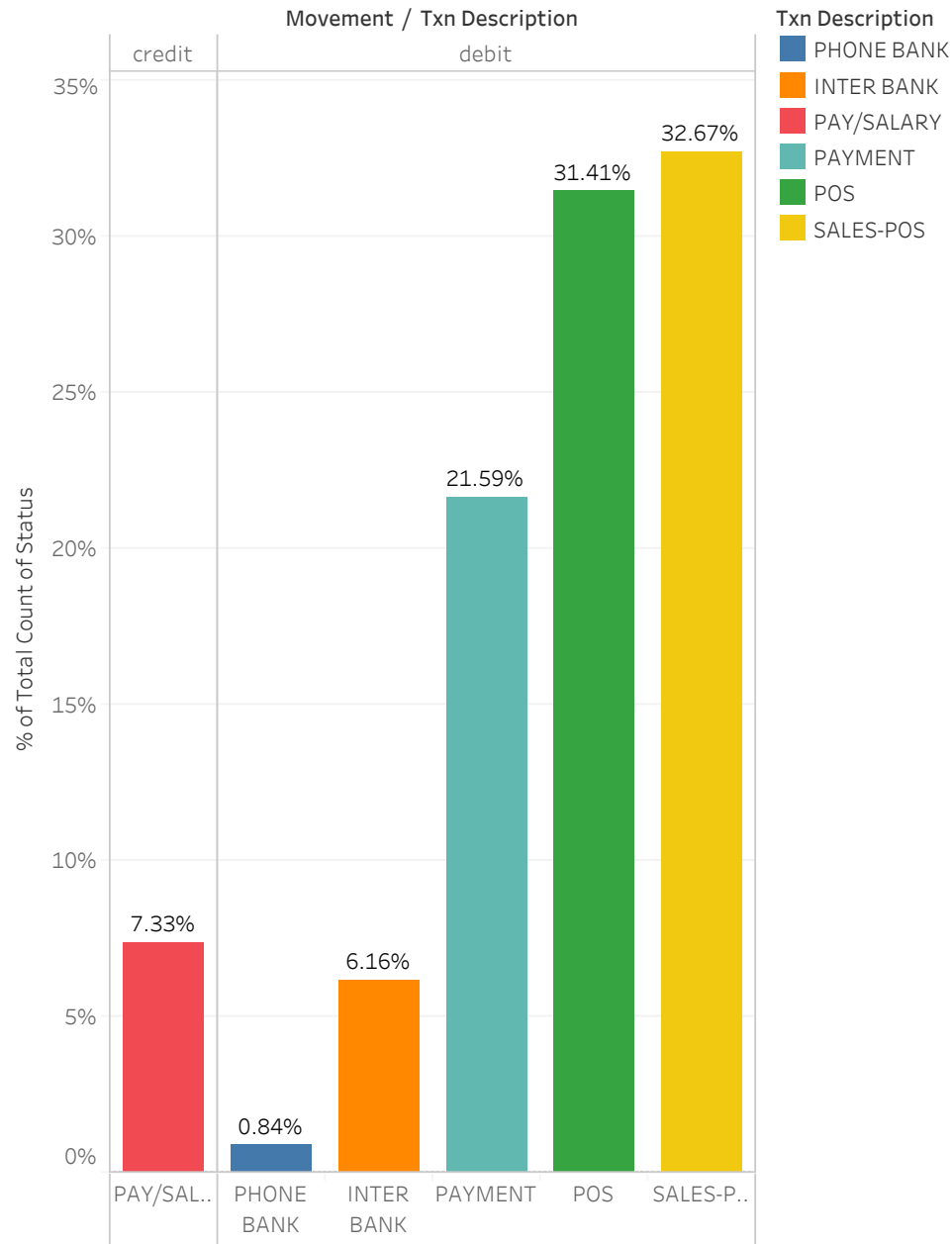
Customer State

Australia

© Mapbox © OSM   Z

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

Customer Distribution By State

State	%
VIC	34.00%
NSW	25.00%
WA	14.00%
QLD	14.00%
SA	7.00%
NT	3.00%
Other	1.00%
ACT	1.00%

Transactions by type

Movement / Txn Description

% of Total Count of Status for each Txn Description broken down by Movement. Color shows details about Txn Description.