ReDI School of Digital Integration

Data Analytics - Fall 2021

Final Project presentation

Sugandh Bansal - Preeti Sharma

Sales & Marketing Dataset (Bike-selling Company)

About the project

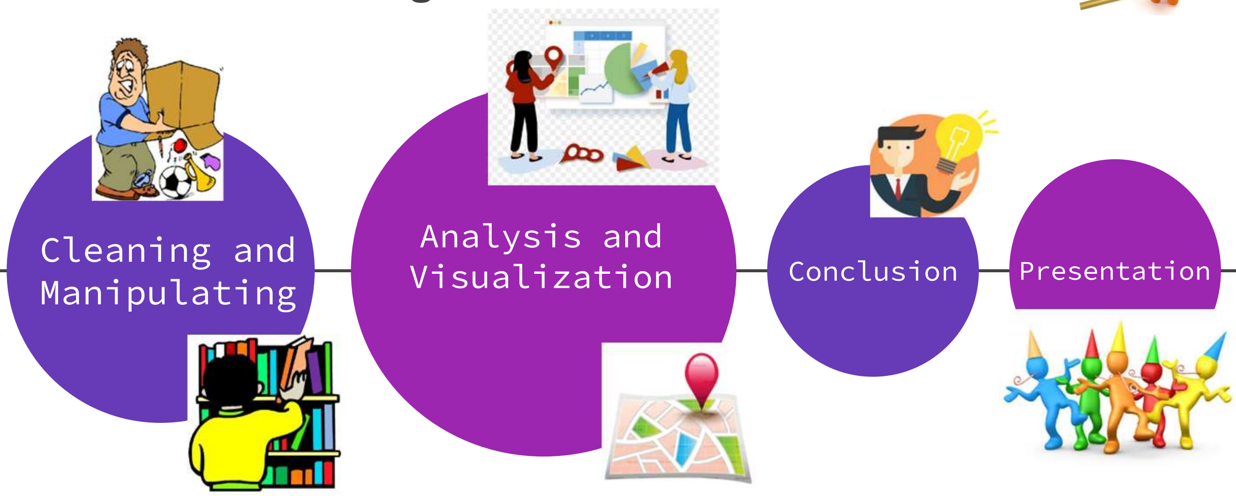
06/12/2021

Dataset – Customer Details, Transaction Details, Customer Demography (One Excel file in different sheets)

Problem – Exploratory Data Analysis – future planning for promotional sales and marketing for existing products

Libraries used – NumPy, Pandas, Matplotlib, Geocoder, Plotly, datetime, missingno

Process and Planning



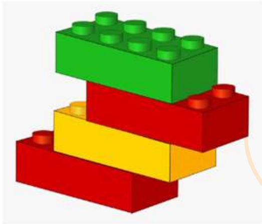
To check duplicay or any null value in the dataframe

```
duplicate_data = df_demo.duplicated()
df_demo[duplicate_data].sum()
```

```
customer_id                            0.0
gender                                 0.0
past_3_years_bike_related_purchases    0.0
job_title                              0.0
job_industry_category                  0.0
wealth_segment                         0.0
owns_car                               0.0
name                                   0.0
age                                    0.0
age_range                              0.0
dtype: float64
```

We'll save merged file to keep the data safe

```
df_demoaddtrans.to_csv('Data/CustomerDemoAddTrans.csv')
```

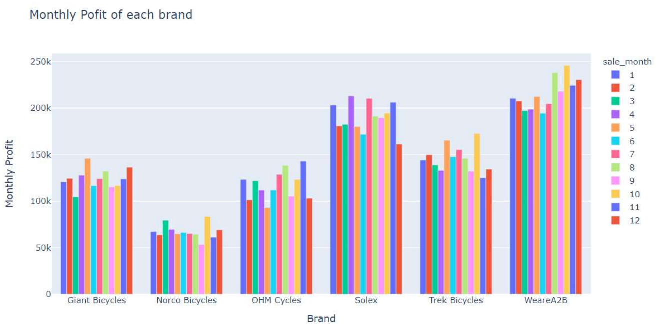


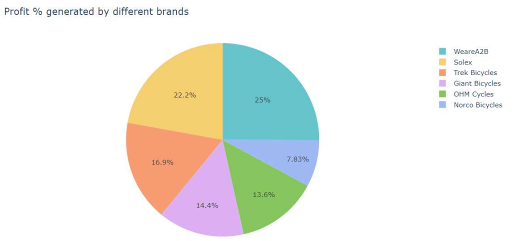
- Problem points to analyze
- What we should visualize and how
- Who will do what part
- Statistical analysis
- Segmentation … and on …

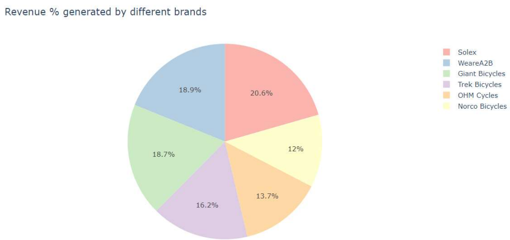
- Brainstorming

Sale of the brand and product line analysis

a) Which brand is most profitable ?





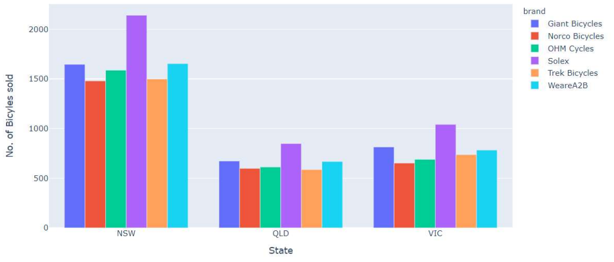


sale_month	1	2	3	4	5	6	7	8	9	10	11	12
brand												
Giant Bicycles	120719.80	124478.99	104599.55	127888.69	146007.33	116562.20	124142.16	132296.42	115150.93	116662.07	123896.37	136474.07
Norco Bicycles	67371.95	63711.91	79455.58	69549.87	64825.82	66283.30	65115.27	64361.79	53358.55	83541.67	61182.56	69074.27
OHM Cycles	123322.74	101334.20	121840.47	111855.60	93196.94	111895.56	128738.02	138325.14	105387.79	123335.93	143031.34	103103.90
Solex	203231.88	180857.54	182480.39	213091.94	180178.39	171810.80	210367.06	191407.14	189655.38	194571.81	206211.75	161324.15
Trek Bicycles	144203.43	149818.50	138915.80	132952.03	165402.02	147687.99	155463.59	146037.22	132219.83	172633.06	125088.25	134355.52
WeareA2B	210533.02	207644.43	197193.32	198802.33	212454.12	194481.27	204685.58	238065.16	218065.05	245764.50	224436.95	230533.37

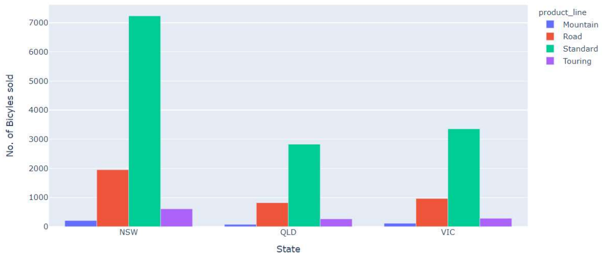
c) How each brand is performing in different states?

brand	Giant Bicycles	Norco Bicycles	OHM Cycles	Solex	Trek Bicycles	WeareA2B
state						
NSW	1646	1479	1587	2140	1499	1653
QLD	673	598	612	848	586	668
VIC	814	652	689	1041	737	783

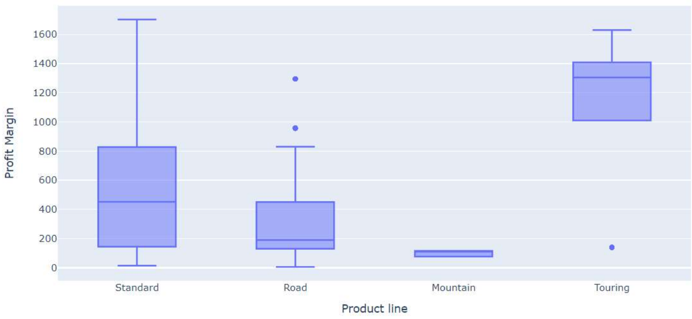
How each brand is performing in different states?



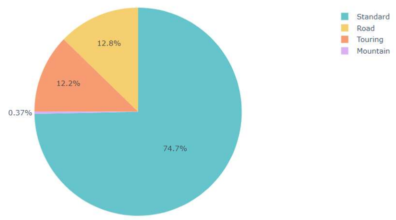
Is any Product-Line preferred over others in different states?



Which Product-line is generating most profit?



Profit % generated by different product-lines



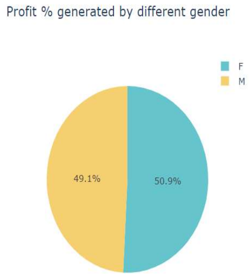
Product-Line sale by Post code

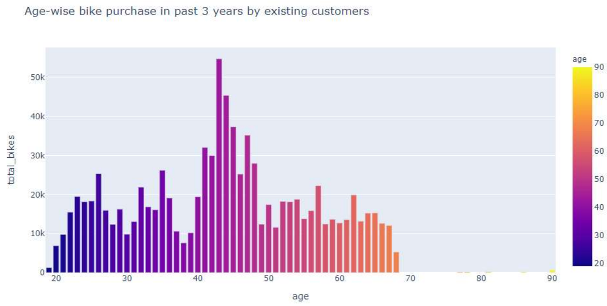
product_line	Mountain	Road	Standard	Touring
postcode				
2000	NaN	10.0	28.0	2.0
2007	NaN	1.0	10.0	2.0
2008	1.0	2.0	3.0	1.0
2009	1.0	6.0	19.0	1.0
2010	4.0	7.0	40.0	4.0

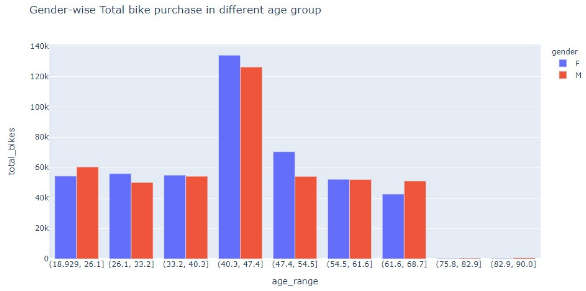
Customer behaviour analysis

a) Which age-group is buying bicycles the most?

gender	F	M
age_range		
(18.929, 26.1]	1080	1149
(26.1, 33.2]	1099	1008
(33.2, 40.3]	1130	1139
(40.3, 47.4]	2722	2551
(47.4, 54.5]	1475	1127
(54.5, 61.6]	1121	1055
(61.6, 68.7]	983	1036
(75.8, 82.9]	3	12
(82.9, 90.0]	5	10



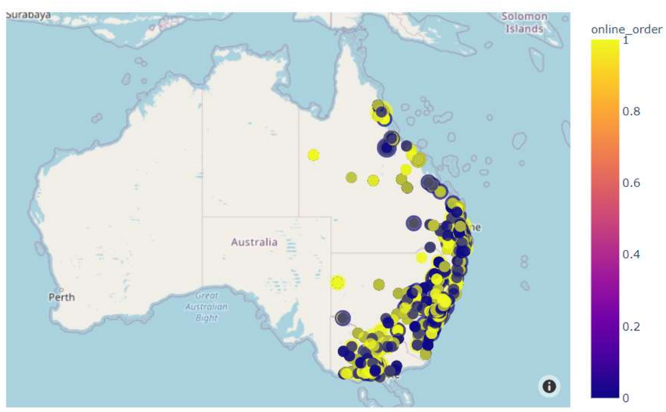




b) Customer's behaviour on purchasing mode i.e online/offline

total transaction count	**18705**
Online/Offline	
0.0	9188
1.0	9517

Online and Offline Transaction count



c) RFM (Recency, Frequency, Monetary) analysis and customer segmentation

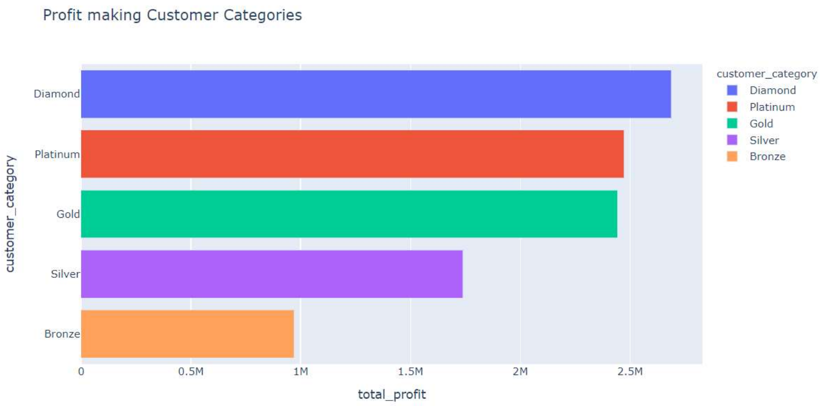
```
rfm_df['customer_category']=pd.qcut(rfm_df['RFM_Score'], 5, ['Diamond','Platinum','Gold','Silver','Bronze'])
rfm_df.sort_values(by='Monetary',ascending=False,inplace=True)
```

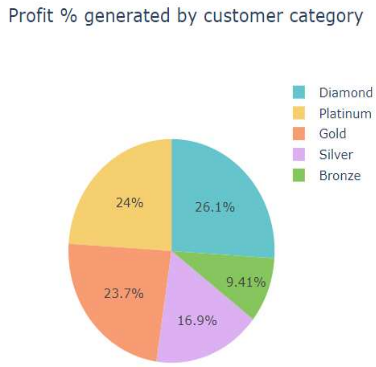
```
# Showing customer category

cust_df = rfm_df.groupby('customer_category').agg({'customer_category':['count'],'Monetary':['sum']}).reset_index()
cust_df.columns=['customer_category','total_count','total_profit']
cust_df
```

	customer_category	total_count	total_profit
0	Diamond	681	2688322.06
1	Platinum	648	2473062.29
2	Gold	757	2443345.96
3	Silver	642	1739129.90
4	Bronze	575	970843.11

```
rfm_df.info()
```

c) Profit generated by segmented customer group



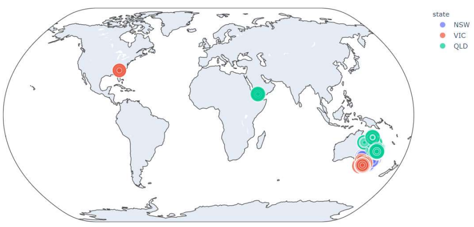


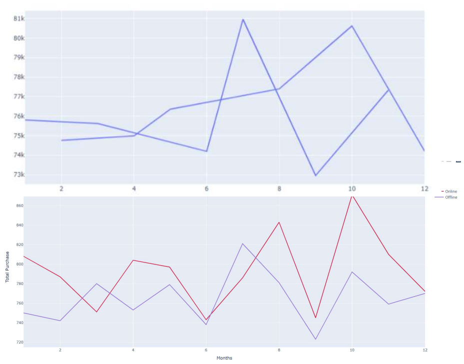
Conclusion

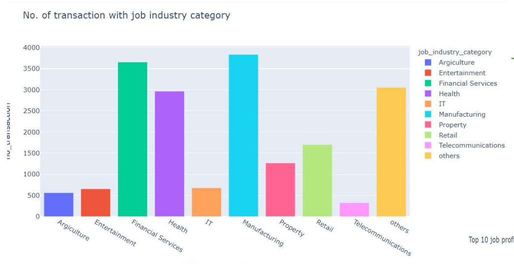
- Target state
- Target Product line
- Target Brand
- Target Selling mode - Online/Offline based on area
- Target segmented group
- Target Age group

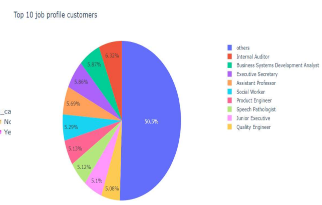
Learned through error: 

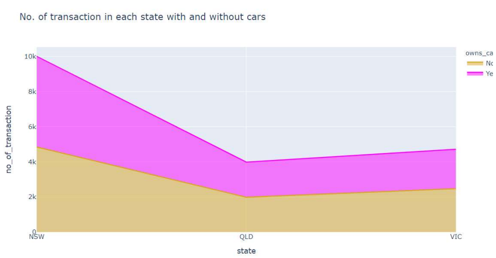
Extra work for fun





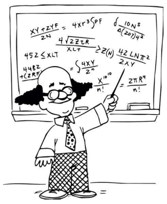






Great Thanks to

Our talented and helpful teachers and ReDI School



TEACHERS

Us