KPMG Data Analytics using python.

The Sporty Rocket Central Dataset

Developing a report that we can present to the client at our next meeting. Displaying the data summary and results of the analysis (see tools/references for assistance). Specifically, the presentation should specify who Sprocket Central Pty Ltd' should be targeting out of the new 1000 customer list using the Transactions Table.

Problem Outline SP Rockect Central is a company that specializes in high quality bikes and accessories. the Company is targeting 1000 new customers and is focused in Maximizing profit through Bike sales.

In [40]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.ticker as mticker
%matplotlib inline
```

In [22]:
```
kpmg = pd.read_csv(r"C:\Users\user\Projects\BI-Analytics-Projects\KPMG Data Analytics (R
```

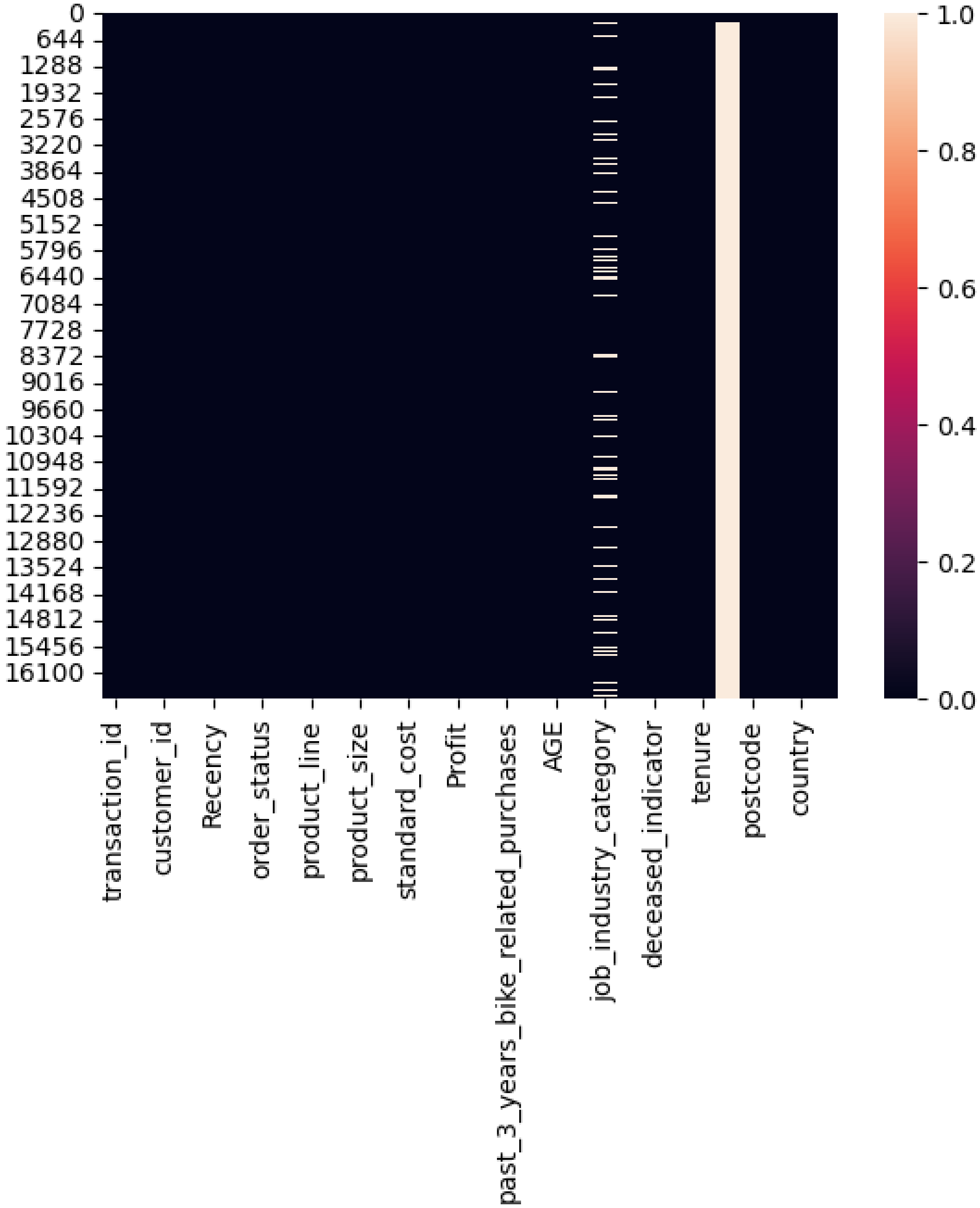
In [3]:
```
kpmg
```

Out[3]:

	transaction_id	product_id	customer_id	transaction_date	Recency	online_order	order_status	brand	pr
0	1	2	2950	2/25/2017	308	False	Approved	Solex	
1	2	3	3120	5/21/2017	223	True	Approved	Trek Bicycles	
2	3	37	402	10/16/2017	75	False	Approved	OHM Cycles	
3	4	88	3135	8/31/2017	121	False	Approved	Norco Bicycles	
4	6	25	2339	3/8/2017	297	True	Approved	Giant Bicycles	
...	...	...	...	...	...	...	...	...	
16715	19995	9	718	5/13/2017	231	True	Approved	OHM Cycles	
16716	19996	51	1018	6/24/2017	189	True	Approved	OHM Cycles	
16717	19997	41	127	11/9/2017	51	True	Approved	Solex	
16718	19998	87	2284	4/14/2017	260	True	Approved	OHM Cycles	
16719	20000	11	1144	9/22/2017	99	True	Approved	Trek Bicycles	

16720 rows × 30 columns

In [4]:
```
kpmg.columns
```
Out[4]:
```
Index(['transaction_id', 'product_id', 'customer_id', 'transaction_date',
       'Recency', 'online_order', 'order_status', 'brand', 'product_line',
       'product_class', 'product_size', 'list_price', 'standard_cost',
       'product_first_sold_date', 'Profit', 'gender',
       'past_3_years_bike_related_purchases', 'DOB', 'AGE', 'job_title',
       'job_industry_category', 'wealth_segment', 'deceased_indicator',
       'owns_car', 'tenure', 'address', 'postcode', 'state', 'country',
       'property_valuation'],
      dtype='object')
```
In [5]:
```
kpmg.dtypes
```
Out[5]:
```
transaction_id                           int64
product_id                               int64
customer_id                              int64
transaction_date                        object
Recency                                  int64
online_order                              bool
order_status                            object
brand                                   object
product_line                            object
product_class                           object
product_size                            object
list_price                             float64
standard_cost                          float64
product_first_sold_date                 object
Profit                                 float64
gender                                  object
past_3_years_bike_related_purchases      int64
DOB                                     object
AGE                                      int64
job_title                               object
job_industry_category                   object
wealth_segment                          object
deceased_indicator                      object
owns_car                                object
tenure                                   int64
address                                 object
postcode                                 int64
state                                   object
country                                 object
property_valuation                       int64
dtype: object
```
In [7]:
```
kpmg.Profit.describe()
```
Out[7]:
```
count    16720.000000
mean       551.874560
std        493.450565
min          4.800000
25%        135.850000
50%        445.210000
75%        827.160000
max       1702.550000
Name: Profit, dtype: float64
```
In [23]:
```
sns.heatmap(kpmg.isnull())
```
Out[23]:
```
<Axes: >
```



In [24]:
```
kpmg.isnull().sum()
```

Out[24]:
```
transaction_id                          0
product_id                              0
customer_id                             0
transaction_date                        0
Recency                                 0
online_order                            0
order_status                            0
brand                                   0
product_line                            0
product_class                           0
product_size                            0
list_price                              0
standard_cost                           0
product_first_sold_date                 0
Profit                                  0
gender                                  0
past_3_years_bike_related_purchases     0
DOB                                     0
AGE                                     0
job_title                               0
job_industry_category                2663
```

```
wealth_segment                              0
deceased_indicator                          0
owns_car                                    0
tenure                                      0
address                                 16513
postcode                                    0
state                                       0
country                                     0
property_valuation                          0
dtype: int64
```

In [68]:
```python
kpmg.drop(columns= ['address'], axis=1, inplace=True)
```

In [26]:
```python
kpmg.isnull().sum()
```

Out[26]:
```
transaction_id                             0
product_id                                 0
customer_id                                0
transaction_date                           0
Recency                                    0
online_order                               0
order_status                               0
brand                                      0
product_line                               0
product_class                              0
product_size                               0
list_price                                 0
standard_cost                              0
product_first_sold_date                    0
Profit                                     0
gender                                     0
past_3_years_bike_related_purchases        0
DOB                                        0
AGE                                        0
job_title                                  0
job_industry_category                   2663
wealth_segment                             0
deceased_indicator                         0
owns_car                                   0
tenure                                     0
postcode                                   0
state                                      0
country                                    0
property_valuation                         0
dtype: int64
```

In [27]:
```python
kpmg.isnull().sum()
```

Out[27]:
```
transaction_id                             0
product_id                                 0
customer_id                                0
transaction_date                           0
Recency                                    0
online_order                               0
order_status                               0
brand                                      0
product_line                               0
product_class                              0
product_size                               0
list_price                                 0
standard_cost                              0
product_first_sold_date                    0
Profit                                     0
gender                                     0
past_3_years_bike_related_purchases        0
DOB                                        0
```

```
AGE                                0
job_title                          0
job_industry_category           2663
wealth_segment                     0
deceased_indicator                 0
owns_car                           0
tenure                             0
postcode                           0
state                              0
country                            0
property_valuation                 0
dtype: int64
```

In [28]:
```python
kpmg.AGE.describe()
```

Out[28]:
```
count    16720.000000
mean        46.192045
std         12.592445
min         22.000000
25%         37.000000
50%         46.000000
75%         55.250000
max         92.000000
Name: AGE, dtype: float64
```

In [29]:
```python
# Group the AGE in bins of 10 years
labels = ["{0} - {1}".format(i, i + 9) for i in range(1, 70, 10)]

kpmg['age_group'] = pd.cut(kpmg.AGE, range(1, 80, 10), right=False, labels=labels)
```

In [30]:
```python
kpmg.age_group.value_counts()
```

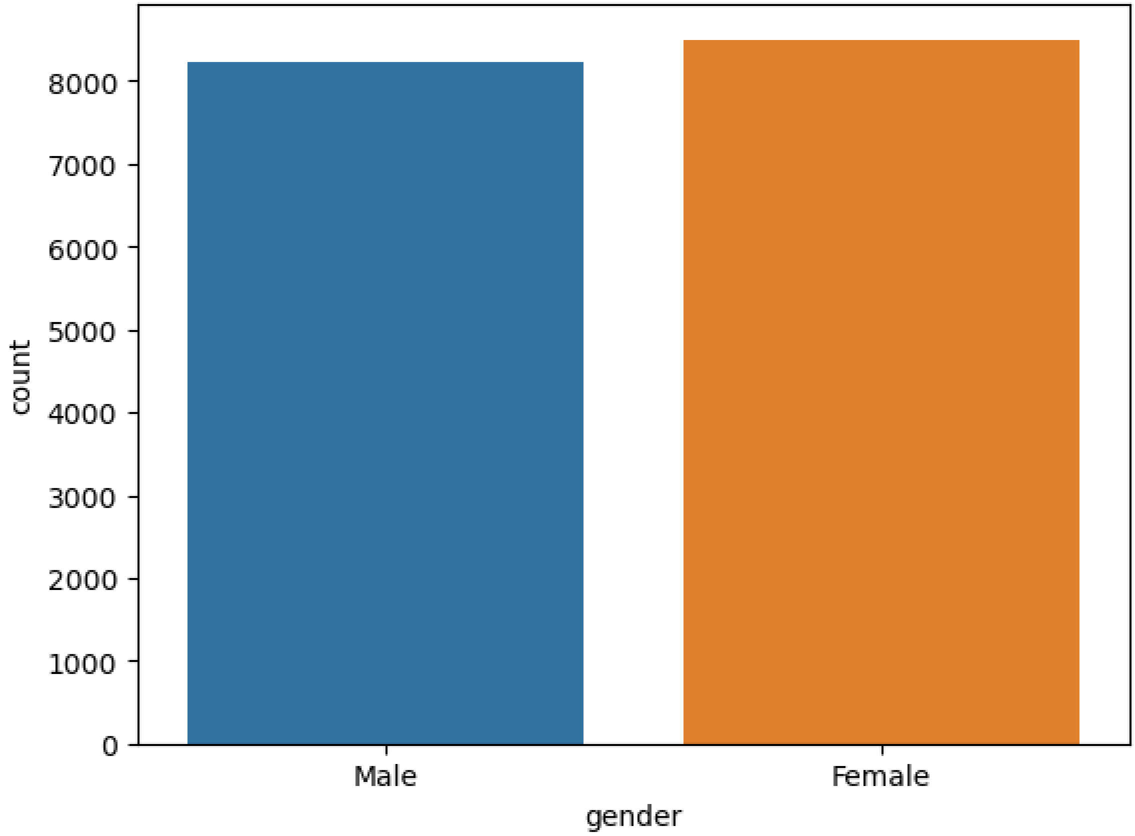
Out[30]:
```
41 - 50    5749
51 - 60    2975
31 - 40    2888
61 - 70    2634
21 - 30    2447
1 - 10        0
11 - 20       0
Name: age_group, dtype: int64
```

In [19]:
```python
kpmg.columns
```

Out[19]:
```
Index(['transaction_id', 'product_id', 'customer_id', 'transaction_date',
       'Recency', 'online_order', 'order_status', 'brand', 'product_line',
       'product_class', 'product_size', 'list_price', 'standard_cost',
       'product_first_sold_date', 'Profit', 'gender',
       'past_3_years_bike_related_purchases', 'DOB', 'AGE', 'job_title',
       'job_industry_category', 'wealth_segment', 'deceased_indicator',
       'owns_car', 'tenure', 'address', 'postcode', 'state', 'country',
       'property_valuation', 'age_group'],
      dtype='object')
```

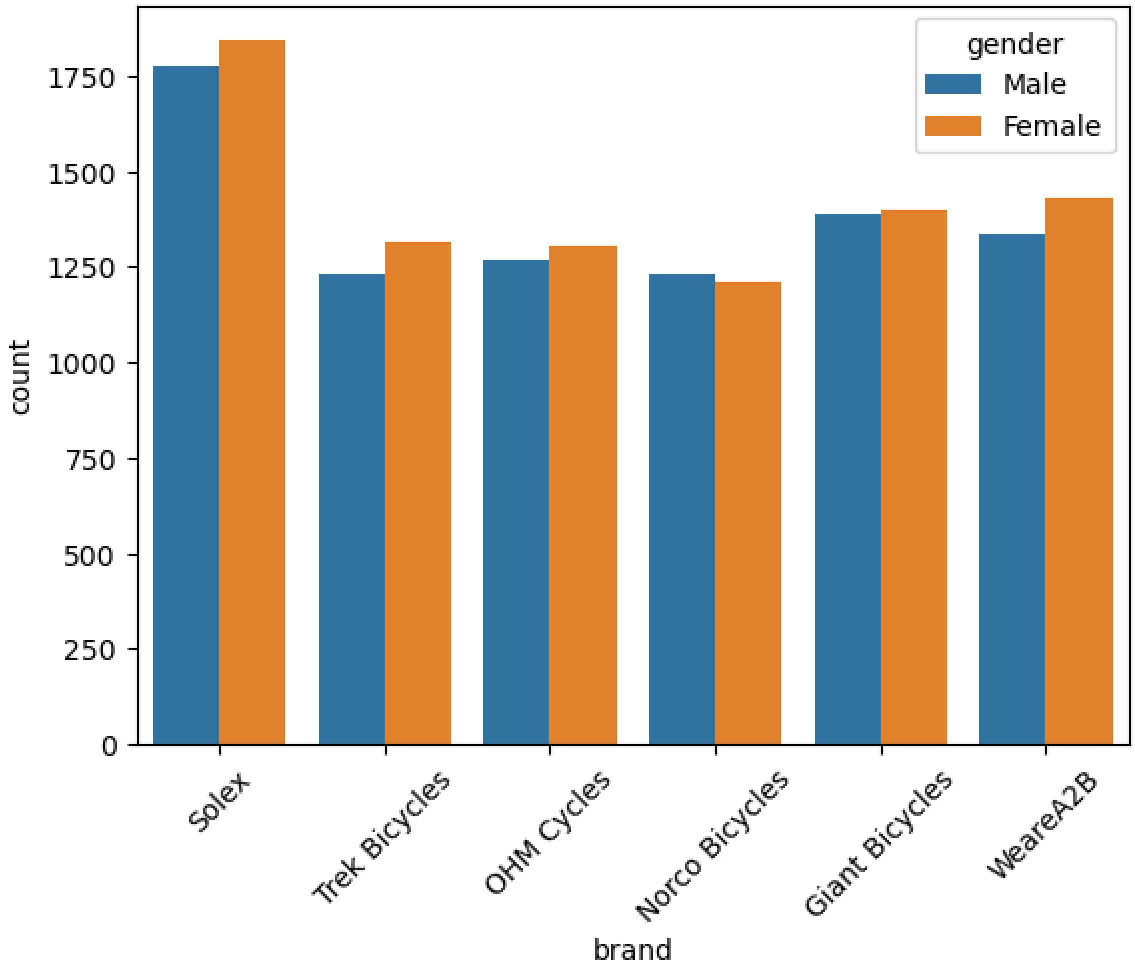
In [31]:
```python
sns.countplot(kpmg, x = 'gender')
```

Out[31]:
```
<Axes: xlabel='gender', ylabel='count'>
```



In [34]:
```
sns.countplot(kpmg, x = 'brand', hue = 'gender')
plt.xticks(rotation=45)
```

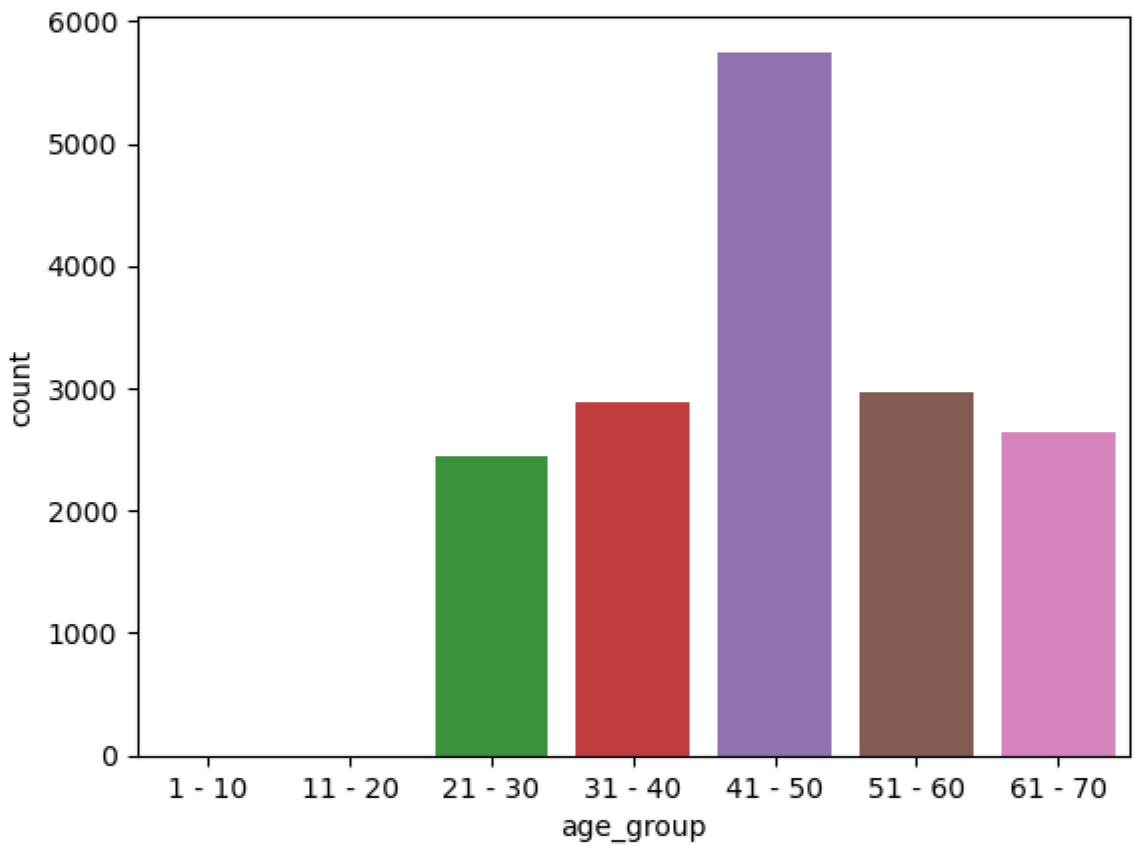
Out[34]:
```
(array([0, 1, 2, 3, 4, 5]),
 [Text(0, 0, 'Solex'),
  Text(1, 0, 'Trek Bicycles'),
  Text(2, 0, 'OHM Cycles'),
  Text(3, 0, 'Norco Bicycles'),
  Text(4, 0, 'Giant Bicycles'),
  Text(5, 0, 'WeareA2B')])
```



In [50]:
```
plt.close()
```

In [53]:
```
sns.countplot(kpmg, x = 'age_group')
```

Out[53]: <Axes: xlabel='age_group', ylabel='count'>



In [65]:
```
kpmg['Profitss'] = kpmg.Profit.sum()
```

In [69]:
```
kpmg
```

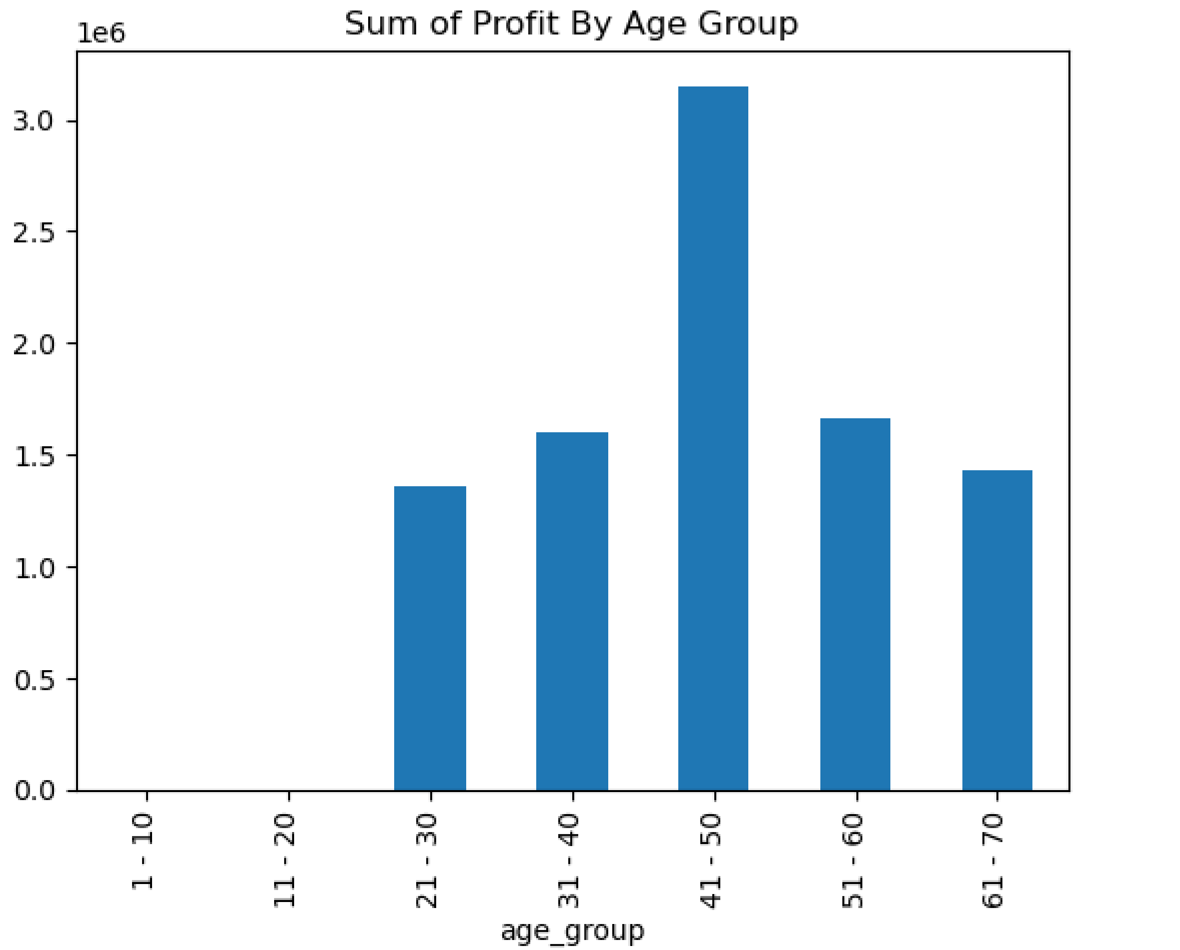
Out[69]:

	transaction_id	product_id	customer_id	transaction_date	Recency	online_order	order_status	brand	pr
0	1	2	2950	2/25/2017	308	False	Approved	Solex	
1	2	3	3120	5/21/2017	223	True	Approved	Trek Bicycles	
2	3	37	402	10/16/2017	75	False	Approved	OHM Cycles	
3	4	88	3135	8/31/2017	121	False	Approved	Norco Bicycles	
4	6	25	2339	3/8/2017	297	True	Approved	Giant Bicycles	
...	...	...	...	...	...	...	...	...	
16715	19995	9	718	5/13/2017	231	True	Approved	OHM Cycles	
16716	19996	51	1018	6/24/2017	189	True	Approved	OHM Cycles	
16717	19997	41	127	11/9/2017	51	True	Approved	Solex	
16718	19998	87	2284	4/14/2017	260	True	Approved	OHM Cycles	
16719	20000	11	1144	9/22/2017	99	True	Approved	Trek Bicycles	

16720 rows × 30 columns

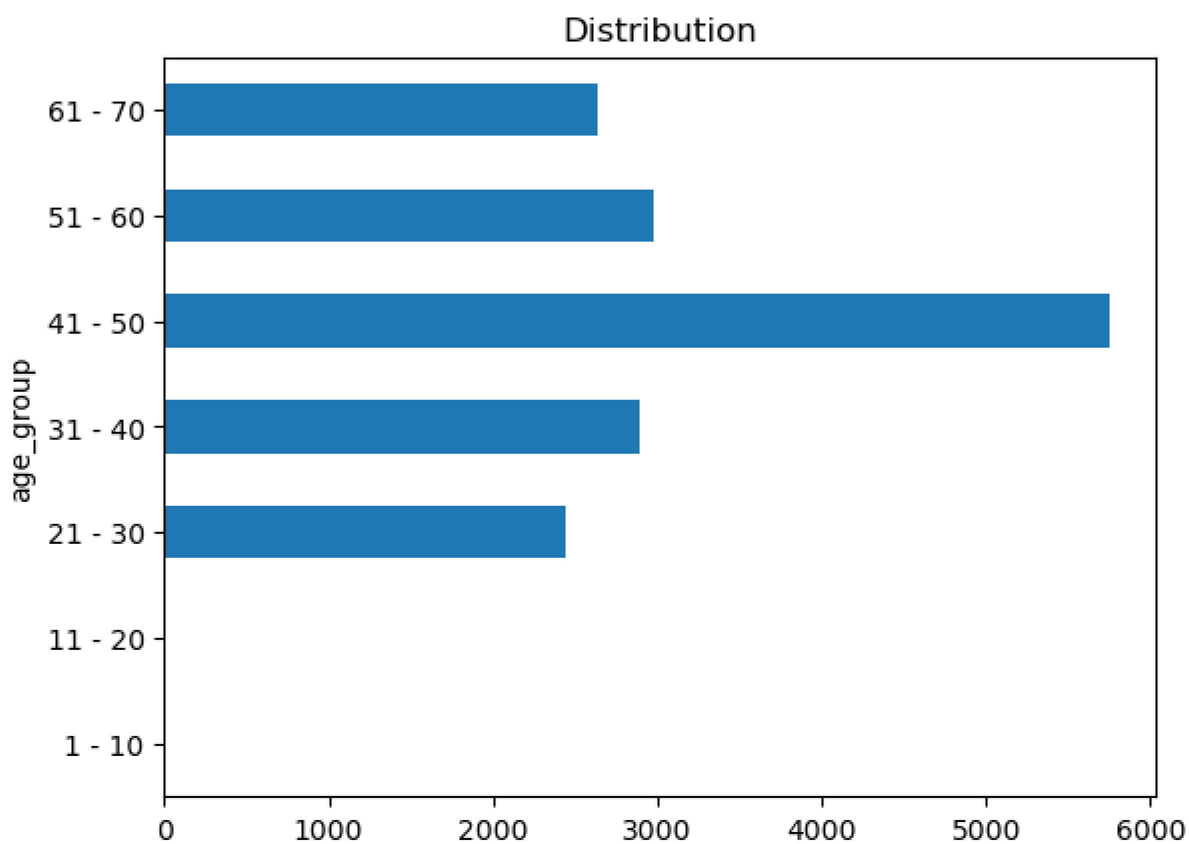
In [83]:
```
kpmg.groupby(kpmg.age_group).Profit.sum().plot(kind = 'bar')
plt.title("Sum of Profit By Age Group")
```

Out[83]:
```
Text(0.5, 1.0, 'Sum of Profit By Age Group')
```



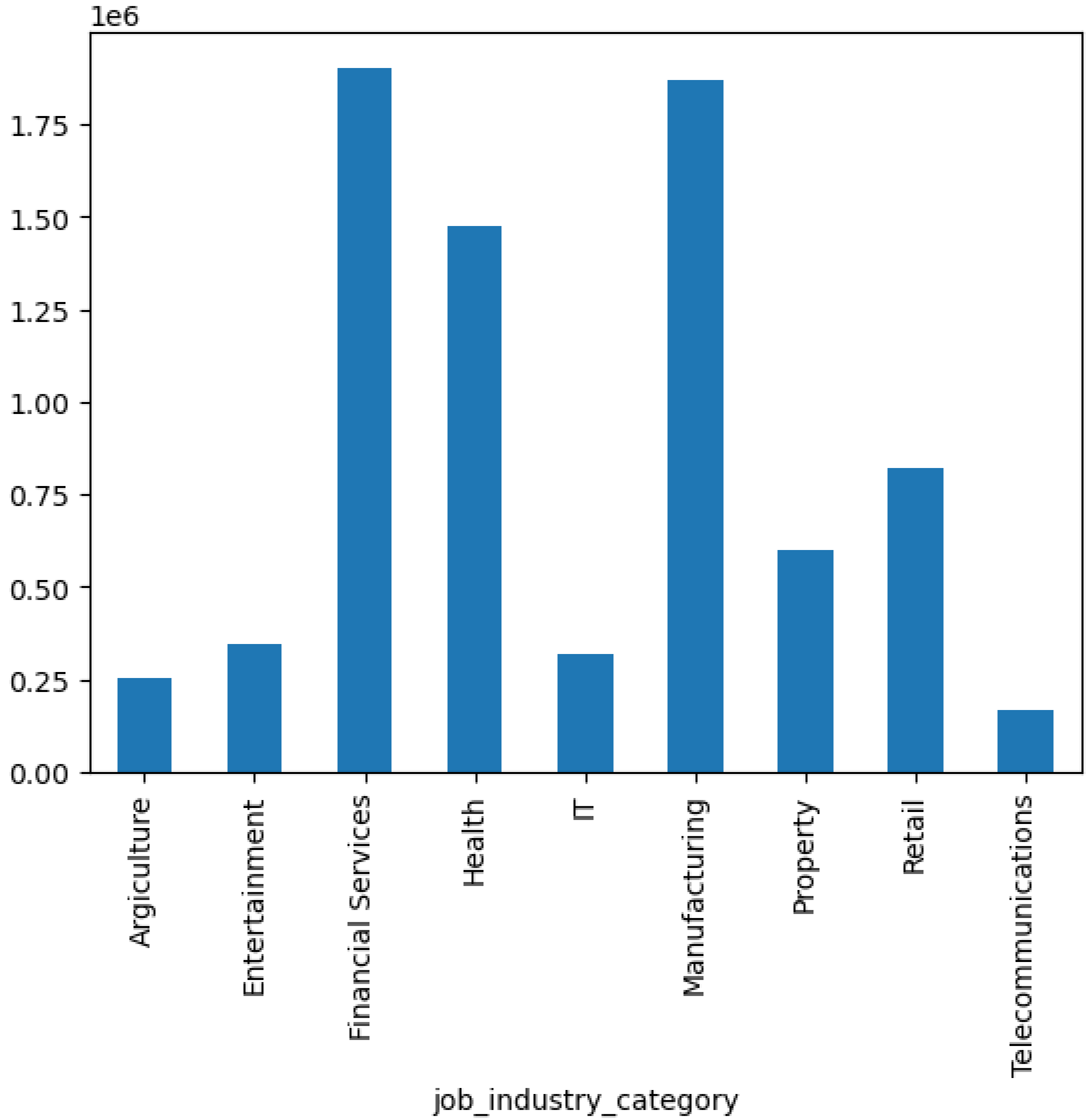
In [81]:
```
kpmg.groupby(kpmg.age_group).past_3_years_bike_related_purchases.count().plot(kind = 'ba
plt.title("Distribution")
```

Out[81]:
```
Text(0.5, 1.0, 'Distribution')
```



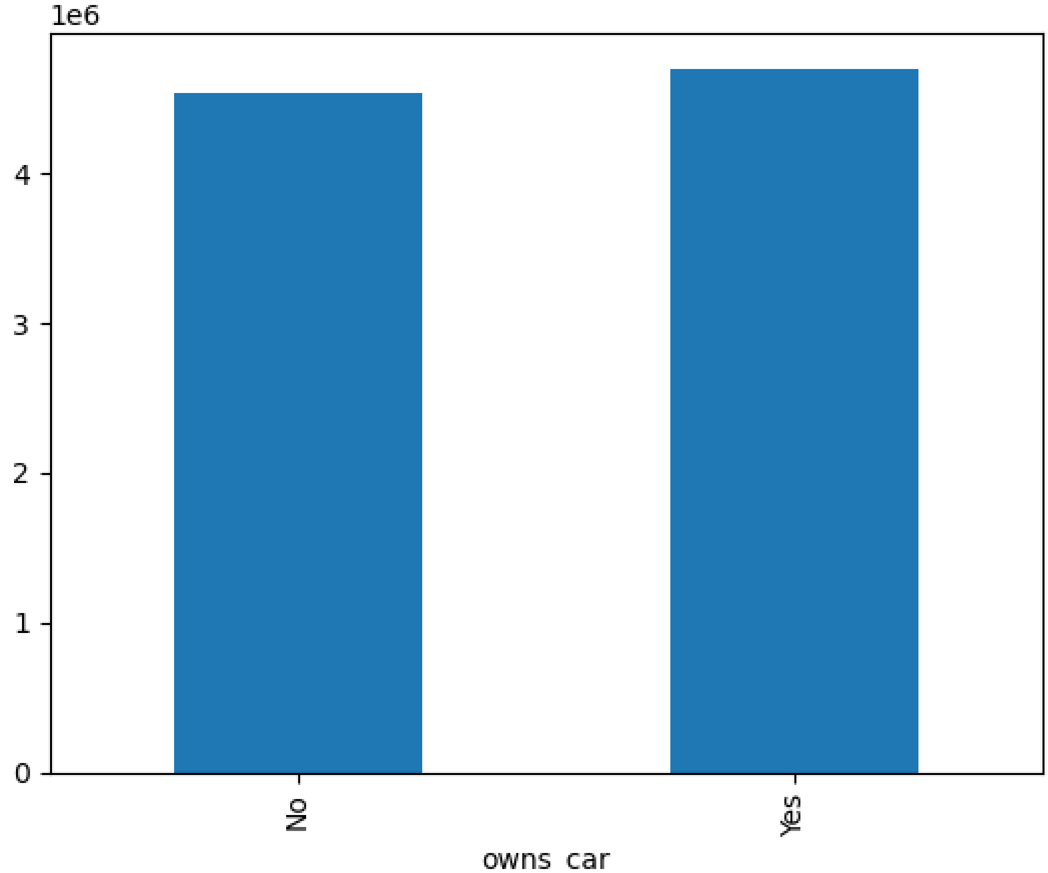
In [82]:
```
kpmg.groupby(kpmg.job_industry_category).Profit.sum().plot(kind = 'bar')
```

Out[82]: <Axes: xlabel='job_industry_category'>



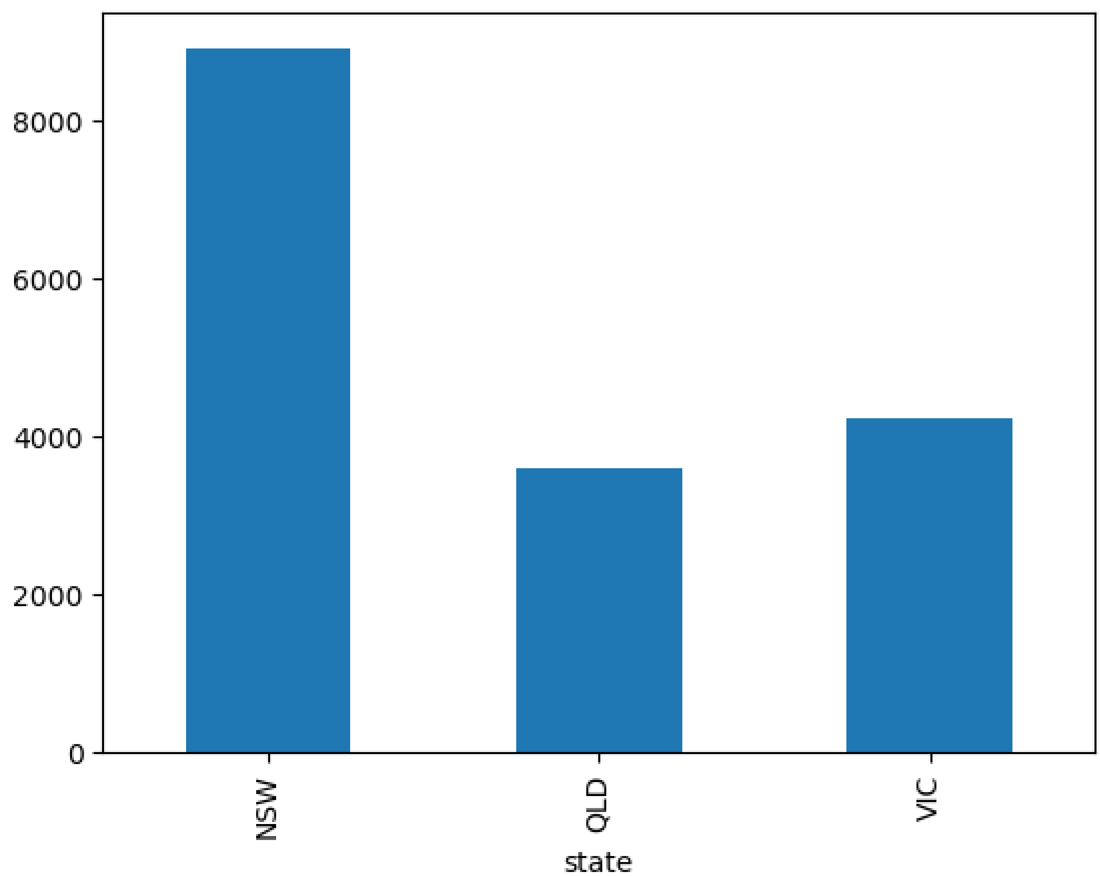
In [84]:
```
kpmg.groupby(kpmg.owns_car).Profit.sum().plot(kind = 'bar')
```

Out[84]: <Axes: xlabel='owns_car'>



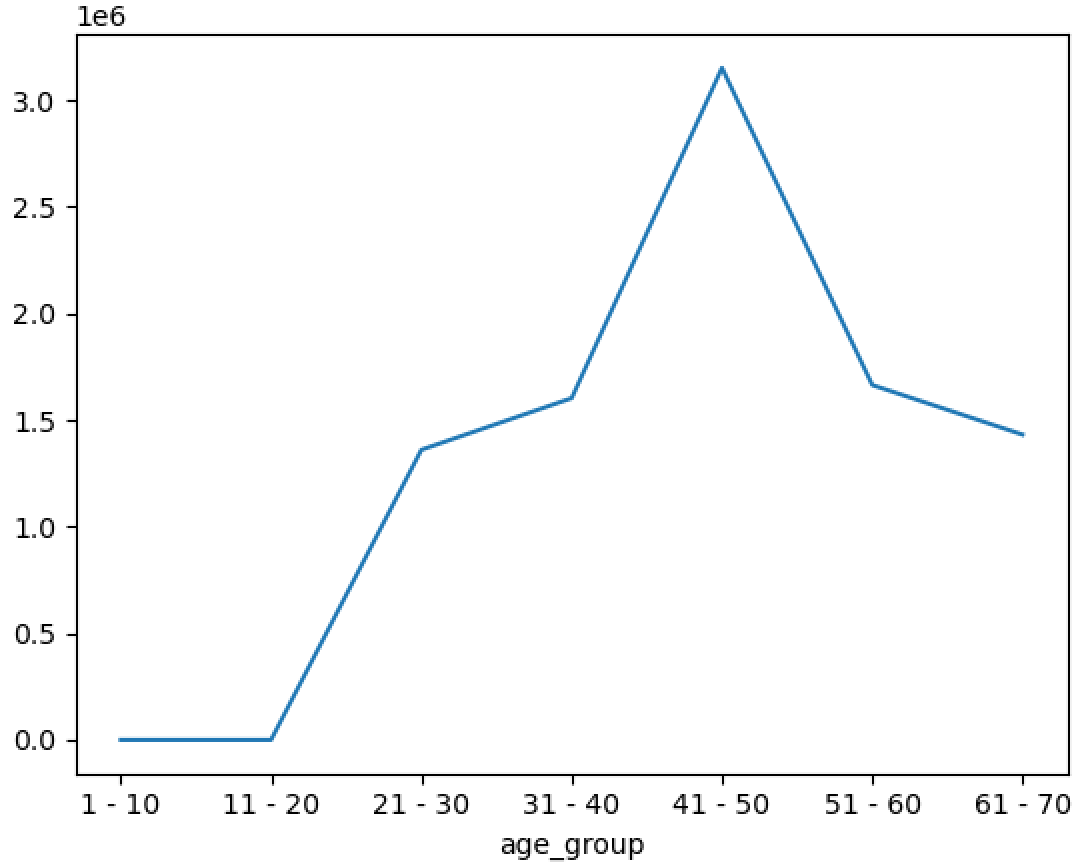
In [85]:
```
kpmg.groupby(kpmg.state).customer_id.count().plot(kind = 'bar')
```

Out[85]: <Axes: xlabel='state'>

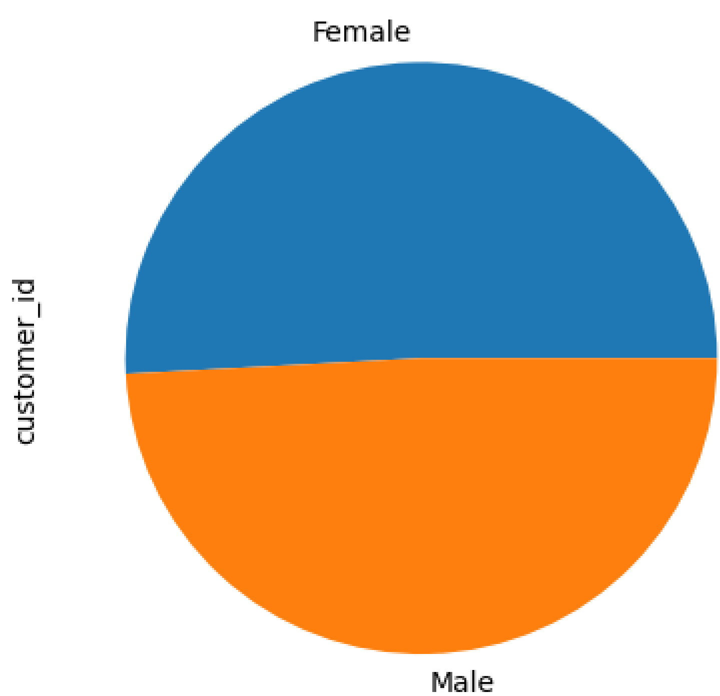


In [86]:
```
kpmg.groupby(kpmg.age_group).Profit.sum().plot(kind = 'line')
```

Out[86]: <Axes: xlabel='age_group'>



In [87]:
```
kpmg.groupby(kpmg.gender).customer_id.count().plot(kind = 'pie')
```
Out[87]:
```
<Axes: ylabel='customer_id'>
```



In [88]:
```
kpmg.gender.value_counts()
```
Out[88]:
```
Female    8501
Male      8219
Name: gender, dtype: int64
```

DATA EXPLORATION

- Most of the Bike related Purchases were made by the age 40 and 49
- The Data shows that middle aged customers are the most potential Customers
- Finalcial Services, Health and Manufacturing Sector are the top Three Profit Generating industries, followed by Retail, IT and Property

In [ ]:

```

```

In [ ]:

```

```