**Technical Test Answers**

Nama : Jafar Shodiq

File : ipynb (terlampir)

**Dataset**

1. Mengkategorikan semua jkt-hub menjadi Jakarta

for i in df['kota']:

if 'hub' in i:

df['kota'] = df['kota'].str.replace(i, 'jakarta')

if 'jakarta' in i:

x = i[:7]

df['kota'] = df['kota'].str.replace(i, x)

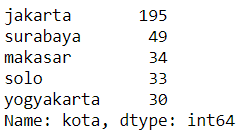
df[['kota', 'status']][df['kota'].str.contains('jakarta')].sample(10)



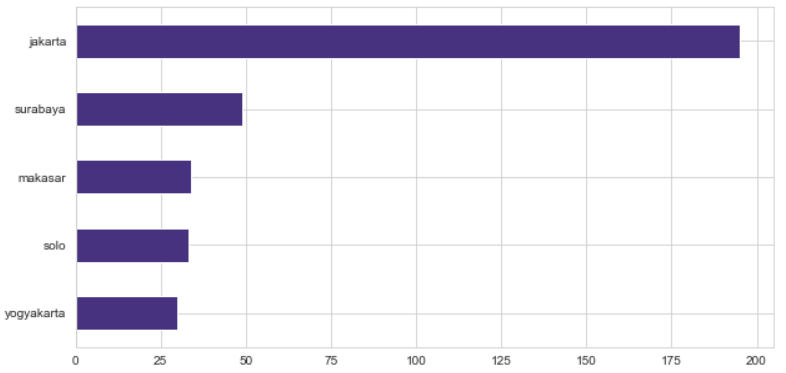
1. Jumlah ATM yang aktif di top 5 kota

df2 = df[['kota', 'status']]

df2['kota'][df2['status'].str.contains('aktif')].value\_counts().head()



df2['kota'][df2['status'].str.contains('aktif')].value\_counts().head().sort\_values().plot(kind='barh', figsize=(10, 5))



1. Di kota Jakarta, penyewaan ATM yang aktif ada di kategori lokasi mana saja

df3 = df[['lokasi']][(df['kota']=='jakarta') & (df['status']=='aktif')]

Me-list semua kategori yang mungkin:

* Swalayan

lst\_swalayan = []

lst\_swalayan\_cat = ['toko', 'swalayan', 'tip top', 'carefour', 'indomaret', 'alfa mart', 'alfamart', 'giant', 'super indo']

for i in lst\_swalayan\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_swalayan.append(x)

len\_swalayan = len(list(set(lst\_swalayan)))

* SPBU

lst\_spbu = []

lst\_spbu\_cat = ['spbu']

for i in lst\_spbu\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_spbu.append(x)

len\_spbu = len(list(set(lst\_spbu)))

* Kantor Cabang Pembantu (KCP)

lst\_kcp = []

lst\_kcp\_cat = ['kcp']

for i in lst\_kcp\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_kcp.append(x)

len\_kcp = len(list(set(lst\_kcp)))

* Mall

lst\_mall = []

lst\_mall\_cat = ['mall', 'plaza']

for i in lst\_mall\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_mall.append(x)

len\_mall = len(list(set(lst\_mall)))

* Universitas

lst\_univ = []

lst\_univ\_cat = ['universitas', 'univ', 'kampus']

for i in lst\_univ\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_univ.append(x)

len\_univ = len(list(set(lst\_univ)))

* PLN

lst\_pln = []

lst\_pln\_cat = ['pln']

for i in lst\_pln\_cat:

for x in list(df3['lokasi'][df3['lokasi'].str.contains(i)]):

lst\_pln.append(x)

len\_pln = len(list(set(lst\_pln)))

Visualisasi:

new\_df\_lokasi = pd.DataFrame(columns=('Cat', 'Count'))

cat\_lst = ['Swalayan', 'Mall', 'SPBU', 'Universitas', 'KCP', 'PLN']

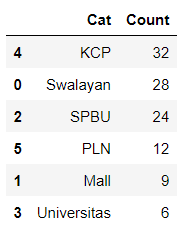
len\_lst = [len\_swalayan, len\_mall, len\_spbu, len\_univ, len\_kcp, len\_pln]

new\_df\_lokasi['Cat'] = cat\_lst

new\_df\_lokasi['Count'] = len\_lst

new\_df\_lokasi.sort\_values('Count', ascending=False, inplace=True)

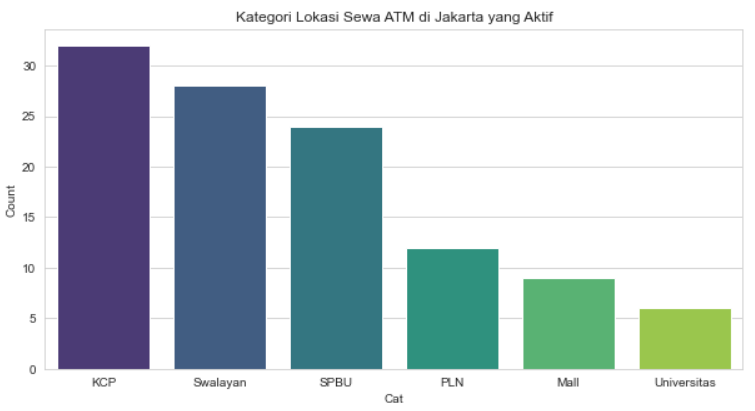
new\_df\_lokasi



plt.figure(figsize=(10, 5))

sns.barplot(x='Cat', y='Count', data=new\_df\_lokasi)

plt.title('Kategori Lokasi Sewa ATM di Jakarta yang Aktif')



1. Korelasi sheet SEWAATM dan Data Aktifitas

data2 = pd.read\_excel('../data-cleaned.xlsx', sheet\_name='data\_aktifitas', index\_col=0)

df4 = data2['lokasi']

df5 = df[['lokasi']][df['status']=='aktif']

lst\_swalayan\_da = []

lst\_swalayan\_sa = []

lst\_swalayan\_cat = ['toko', 'swalayan', 'tip top', 'carefour', 'indomaret', 'alfa mart', 'alfamart', 'giant', 'super indo']

for i in lst\_swalayan\_cat:

for x in list(df4[df4.str.contains(i)]):

lst\_swalayan\_da.append(x)

for x in list(df5['lokasi'][df5['lokasi'].str.contains(i)]):

lst\_swalayan\_sa.append(x)

len\_swalayan\_da = len(list(set(lst\_swalayan\_da)))

len\_swalayan\_sa = len(list(set(lst\_swalayan\_sa)))

new\_df = pd.DataFrame(columns=('cat', 'count'))

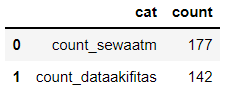
cat\_lst = ['count\_sewaatm', 'count\_dataakifitas']

len\_lst = [len\_swalayan\_sa, len\_swalayan\_da]

new\_df['cat'] = cat\_lst

new\_df['count'] = len\_lst

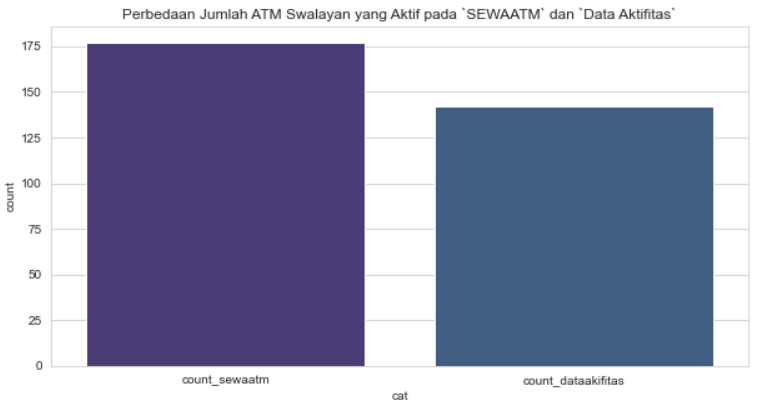
new\_df



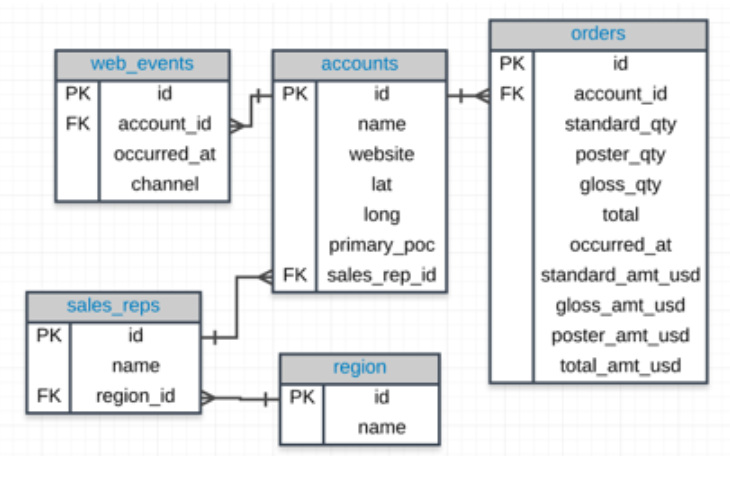
plt.figure(figsize=(10, 5))

sns.barplot(x='cat', y='count', data=new\_df)

plt.title('Perbedaan Jumlah ATM Swalayan yang Aktif pada `SEWAATM` dan `Data Aktifitas`')



**SQL**



1. Provide the **name** for each region for every **order**, as well as the account **name** and the **unit price** they paid (total\_amt\_usd/total) for the order. However, you should only provide the results if the **standard order quantity** exceeds 100 and the **poster order quantity** exceeds 50. Your final table should have 3 columns: **region name**, **account name**, and **unit price**. Sort for the largest **unit price** first.

SELECT region.name AS region\_name, accounts.name AS account\_name, (orders.total\_amt\_usd/orders.total) AS unit\_price

FROM accounts

JOIN orders

ON orders.account\_id=accounts.id

JOIN sales\_reps

ON sales\_reps.id=accounts.sales\_rep\_id

JOIN region

ON region.id=sales\_reps.region\_id

WHERE orders.standard\_qty>100 AND orders.poster\_qty>50

ORDER BY unit\_price DESC

1. Provide the **name** of the **sales\_rep** in each **region** with the largest amount of **total\_amt\_usd** sales.

SELECT sales\_reps.name, MAX(orders.total\_amt\_usd)

FROM sales\_reps

JOIN accounts

ON sales\_reps.id=accounts.sales\_rep\_id

JOIN orders

ON orders.account\_id=accounts.id

WHERE sales\_reps.region\_id=1

UNION

SELECT sales\_reps.name, MAX(orders.total\_amt\_usd)

FROM sales\_reps

JOIN accounts

ON sales\_reps.id=accounts.sales\_rep\_id

JOIN orders

ON orders.account\_id=accounts.id

WHERE sales\_reps.region\_id=2

1. Find all the orders that occurred in 2015. Your final table should have 4 columns: **occurred\_at**, **account name**, **order total**, and **order total\_amt\_usd**.

SELECT orders.occured\_at, accounts.name, orders.total, orders.total\_amt\_usd

FROM orders

JOIN accounts

ON orders.account\_id=accounts.id

WHERE orders.occured\_at=2015