Dataset Description

https://github.com/Machinfy/Data-Analysis-with-python-Crash-Course/blob/main/Orange_Telecom_Churn_Data.csv

data dicitionary, or description of the variables:

- **state** : the state code for this user (String)
- **account_length** : the age of this account - consider it into months - (int)
- **area_code** : the code for area for this account (int)
- **phone_number** : the account number for this user (string)
- **intl_plan** : if user has intl plan or not (string)
- **voice_mail_plan** : if user has voice mail plan or not (string)
- **number_vmail_messages** : count of voice mail messages sent by this user (int)
- **total_day_minutes** : the total minutes for day (float)
- **total_day_calls** : count of calls made at day (int)
- **total_day_charge** : cost of all minuts and calls during day (float)
- **total_eve_minutes** : the total minutes for eve (float)
- **total_eve_calls** : count of calls made at eve (int)
- **total_eve_charge** : cost of all minuts and calls during eve (float)
- **total_night_minutes** : the total minutes for night (float)
- **total_night_calls** : count of calls made at night (int)
- **total_night_charge** : cost of all minuts and calls during night (float)
- **total_intl_minutes** : the total minutes for intl (float)
- **total_intl_calls** : the total calls made for intl (int)
- **total_intl_charge** : cost of all minuts and calls during intl (float)
- **number_customer_service_calls** : how many times he called customer service (int)
- **churned** : churned or not (boolean)

Business Questions

- how many states we have and what is the count for users in each state?
- what is the age of the oldest account we have in our company?
- how many intl plan subscriber?
- how many voice mail subscriber?
- what is the average voice mail messages sent by who subscribed in voice mail plan **ONLY**?
- what is the average minutes for day, eve, night and intl?
- what is the average call counts for day, eve, night and intl?
- what is the average charge for day, eve, night and intl?
- what is the churn rate in our company?
- True of False for:
  - calls during day is more worthy than night?
  - charge during night is cheaper than eve?
- recommend the best time we can offer a free minutes for customers? [optional]

In [1]:

```
import pandas as pd
import numpy  as py
```

In [31]:

```
df = pd.read_csv('Orange_Telecom_Churn_Data.csv')
df.head(10)
```

Out[31]:

	state	account_length	area_code	phone_number	intl_plan	voice_mail_plan	number_vmail_mess
0	KS	128	415	382-4657	no	yes	
1	OH	107	415	371-7191	no	yes	
2	NJ	137	415	358-1921	no	no	
3	OH	84	408	375-9999	yes	no	
4	OK	75	415	330-6626	yes	no	
5	AL	118	510	391-8027	yes	no	
6	MA	121	510	355-9993	no	yes	
7	MO	147	415	329-9001	yes	no	
8	LA	117	408	335-4719	no	no	
9	WV	141	415	330-8173	yes	yes	

10 rows × 21 columns

In [ ]:

```
```

In [3]:

```
df['state'].value_counts()
```

Out[3]:

```
WV    158
MN    125
AL    124
ID    119
VA    118
OH    116
TX    116
WY    115
NY    114
OR    114
NJ    112
UT    112
WI    106
MI    103
ME    103
MA    103
MD    102
VT    101
MT     99
RI     99
KY     99
MS     99
```

```
CT     99
KS     99
WA     98
IN     98
CO     96
NH     95
DE     94
MO     93
AR     92
NM     91
NC     91
SC     91
FL     90
NV     90
OK     90
TN     89
AZ     89
DC     88
IL     88
NE     88
ND     88
HI     86
SD     85
GA     83
LA     82
PA     77
AK     72
IA     69
CA     52
Name: state, dtype: int64
```

In [7]:

```
x= pd.DataFrame(df['state'].value_counts())
x
```

Out[7]:

	state
WV	158
MN	125
AL	124
ID	119
VA	118
OH	116
TX	116
WY	115
NY	114
OR	114
NJ	112
UT	112
WI	106

	state
MI	103
ME	103
MA	103
MD	102
VT	101
MT	99
RI	99
KY	99
MS	99
CT	99
KS	99
WA	98
IN	98
CO	96
NH	95
DE	94
MO	93
AR	92
NM	91
NC	91
SC	91
FL	90
NV	90
OK	90
TN	89
AZ	89
DC	88
IL	88
NE	88
ND	88
HI	86
SD	85
GA	83
LA	82
PA	77

	state
AK	72
IA	69
CA	52

In [9]:
```
x.size
```

Out[9]:
```
51
```

In [11]:
```
x.shape
```

Out[11]:
```
(51, 1)
```

how many states we have and what is the count for users in each state?

In [12]:
```
 print ('we have 51 satate ')
```

```
we have 51 satate
```

In [21]:
```
print("the count for users in each state = ",
      x)
```

```
the count for users in each state =       state
WV    158
MN    125
AL    124
ID    119
VA    118
OH    116
TX    116
WY    115
NY    114
OR    114
NJ    112
UT    112
WI    106
MI    103
ME    103
MA    103
MD    102
VT    101
MT     99
RI     99
KY     99
MS     99
CT     99
KS     99
WA     98
IN     98
```

```
CO     96
NH     95
DE     94
MO     93
AR     92
NM     91
NC     91
SC     91
FL     90
NV     90
OK     90
TN     89
AZ     89
DC     88
IL     88
NE     88
ND     88
HI     86
SD     85
GA     83
LA     82
PA     77
AK     72
IA     69
CA     52
```

what is the age of the oldest account we have in our company?

In [22]:
```
df["account_length"].max()
```

Out[22]:
```
243
```

In [23]:
```
print("the age of the oldest account we have in our company = "
,df["account_length"].max())
```

```
the age of the oldest account we have in our company =  243
```

how many intl plan subscriber

In [25]:
```
df["intl_plan"].value_counts()
```

Out[25]:
```
no     4527
yes     473
Name: intl_plan, dtype: int64
```

In [27]:
```
print('intl plan subscriber = 473')
```

```
intl plan subscriber = 473
```

how many voice mail subscriber?

In [28]:
```
df['voice_mail_plan'].value_counts()
```

Out[28]:
```
no     3677
yes    1323
Name: voice_mail_plan, dtype: int64
```

In [29]:
```
print("voice mail subscriber = 1323")
```

```
voice mail subscriber = 1323
```

In [30]:
```
df.head()
```

Out[30]:

	state	account_length	area_code	phone_number	intl_plan	voice_mail_plan	number_vmail_mess
0	KS	128	415	382-4657	no	yes	
1	OH	107	415	371-7191	no	yes	
2	NJ	137	415	358-1921	no	no	
3	OH	84	408	375-9999	yes	no	
4	OK	75	415	330-6626	yes	no	

5 rows × 21 columns

what is the average voice mail messages sent by who subscribed in voice mail plan ONLY

In [36]:
```
df.groupby(['voice_mail_plan','intl_plan']).mean()
```

Out[36]:

		account_length	area_code	number_vmail_messages	total_day_minutes
voice_mail_plan	**intl_plan**				
no	**no**	100.319340	436.353823	0.000000	179.682969
	yes	101.368421	443.166667	0.000000	185.481871
yes	**no**	99.373322	436.807886	29.254195	180.122735
	yes	103.870229	435.717557	29.809160	183.669466

In [37]:
```
print('the average voice mail messages sent by who subscribed in voice mail plan ONLY = 29.254195')
```

```
the average voice mail messages sent by who subscribed in voice mail plan ONLY = 29.254195
```

what is the average minutes for day, eve, night and intl?

In [42]:
```
df['total_day_minutes'].mean()
```

Out[42]:
```
180.2888999999998
```

In [43]:
```
df['total_eve_minutes'].mean()
```

Out[43]:
```
200.63656000000023
```

In [44]:
```
df['total_night_minutes'].mean()
```

Out[44]:
```
200.3916200000002
```

In [45]:
```
df['total_intl_minutes'].mean()
```

Out[45]:
```
10.261779999999993
```

what is the average call counts for day, eve, night and intl?

In [47]:
```
print('average call counts for day, eve, night and intl = 100.029400
,100.19100 ,    99.919200  ,4.435200          ')
```

```
average call counts for day, eve, night and intl = 100.029400 ,100.19100 ,      99.
919200  ,4.435200
```

what is the average charge for day, eve, night and intl?

In [46]:
```
df.describe()
```

Out[46]:

	account_length	area_code	number_vmail_messages	total_day_minutes	total_day_calls	tot
count	5000.00000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	100.25860	436.911400	7.755200	180.288900	100.029400	
std	39.69456	42.209182	13.546393	53.894699	19.831197	
min	1.00000	408.000000	0.000000	0.000000	0.000000	
25%	73.00000	408.000000	0.000000	143.700000	87.000000	
50%	100.00000	415.000000	0.000000	180.100000	100.000000	
75%	127.00000	415.000000	17.000000	216.200000	113.000000	
max	243.00000	510.000000	52.000000	351.500000	165.000000	

In [48]:
```
print(' the average charge for day, eve, night and intl = 30.649668  ,
17.054322 ,9.017732 , 2.771196  ')
```

```
the average charge for day, eve, night and intl = 30.649668  , 17.054322 ,9.017732
, 2.771196
```

what is the churn rate in our company

In [53]:
```
#مش فاهم هنحسب التغير بالنسبه لي ايه فدي هبده مني

df.groupby(['churned']).sum()
```

Out[53]:

churned	account_length	area_code	number_vmail_messages	total_day_minutes	total_day_calls	to
False	428944	1874911	35597	754480.0	428869	
True	72349	309646	3179	146964.5	71278	

True of False for:

calls during day is more worthy than night?

charge during night is cheaper than eve?

In [54]:
```
df.head()
```

Out[54]:

	state	account_length	area_code	phone_number	intl_plan	voice_mail_plan	number_vmail_mess
0	KS	128	415	382-4657	no	yes	
1	OH	107	415	371-7191	no	yes	
2	NJ	137	415	358-1921	no	no	
3	OH	84	408	375-9999	yes	no	
4	OK	75	415	330-6626	yes	no	

5 rows × 21 columns

In [55]:
```
df.describe()
```

Out[55]:

	account_length	area_code	number_vmail_messages	total_day_minutes	total_day_calls	tot
count	5000.00000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	100.25860	436.911400	7.755200	180.288900	100.029400	
std	39.69456	42.209182	13.546393	53.894699	19.831197	
min	1.00000	408.000000	0.000000	0.000000	0.000000	
25%	73.00000	408.000000	0.000000	143.700000	87.000000	
50%	100.00000	415.000000	0.000000	180.100000	100.000000	
75%	127.00000	415.000000	17.000000	216.200000	113.000000	
max	243.00000	510.000000	52.000000	351.500000	165.000000	

calls during day is more worthy than night?

no

charge during night is cheaper than eve? yes

recommend the best time we can offer a free minutes for customers? at intl

In [ ]:

In [ ]:

In [ ]: