```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


#Dataset
df = pd.read_csv("/content/train.csv")
df.shape
```

```
(4250, 20)
```

```
df.isna().sum() #checking null values
```

```
state                           0
account_length                  0
area_code                       0
international_plan              0
voice_mail_plan                 0
number_vmail_messages           0
total_day_minutes               0
total_day_calls                 0
total_day_charge                0
total_eve_minutes               0
total_eve_calls                 0
total_eve_charge                0
total_night_minutes             0
total_night_calls               0
total_night_charge              0
total_intl_minutes              0
total_intl_calls                0
total_intl_charge               0
number_customer_service_calls   0
churn                           0
dtype: int64
```

There are not missing values in any features.

```
df.churn.value_counts()
```

```
no     3652
yes     598
Name: churn, dtype: int64
```

Exploring Categorical features

```
columns = df.columns
cat_features = []
```

```
for col in columns:
    if df[col].value_counts().shape[0] == 2: # Binary Categorical Features
        cat_features.append(col)

cat_features
```

```
['international_plan', 'voice_mail_plan', 'churn']
```

Categorical features with Multiple classes

```
#Extract the numerical features from the dataset
num_var = [feature for feature in df.columns if df[feature].dtypes != 'O']
print('List of Numerical featues {}'.format(num_var))
```

```
List of Numerical featues ['account_length', 'number_vmail_messages', 'total_day_minutes
```

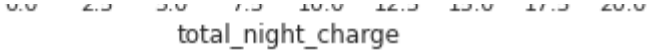
```
#display the all the categorical variable
for feature in num_var:
    sns.set(style = 'whitegrid')
    plt.figure(figsize=(20,5))
    total = len(df)
    ax = sns.countplot(x = df[feature], data = df)
    #plt.title(feature)
    with_per(total, ax)
    plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-35-dcc312951363> in <module>()
      6     ax = sns.countplot(x = df[feature], data = df)
      7     #plt.title(feature)
----> 8     with per(total, ax)
```

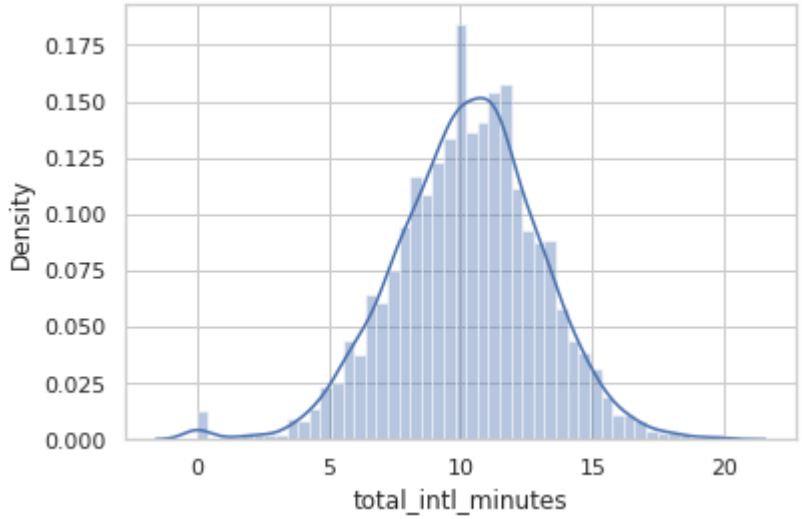
See the above plot 90.7% customers didn't have international plan 73.8% customers didn't have voice mail plan 49.6% customers are living in the area code area_code_415. only 14.1% customers are churn

Univariate Analysis

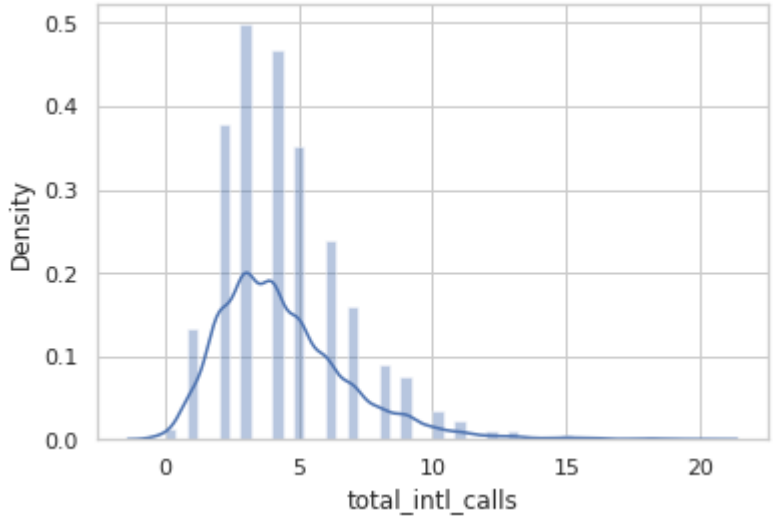

```
# Density plot of all the numerical features
for feature in num_var:
    sns.distplot(df[feature])
    plt.xlabel(feature)
    plt.ylabel('Density')
    plt.show()
```



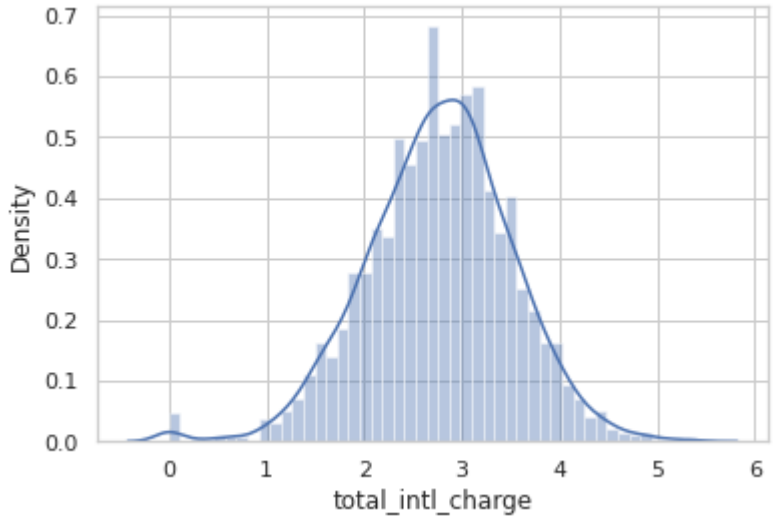

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



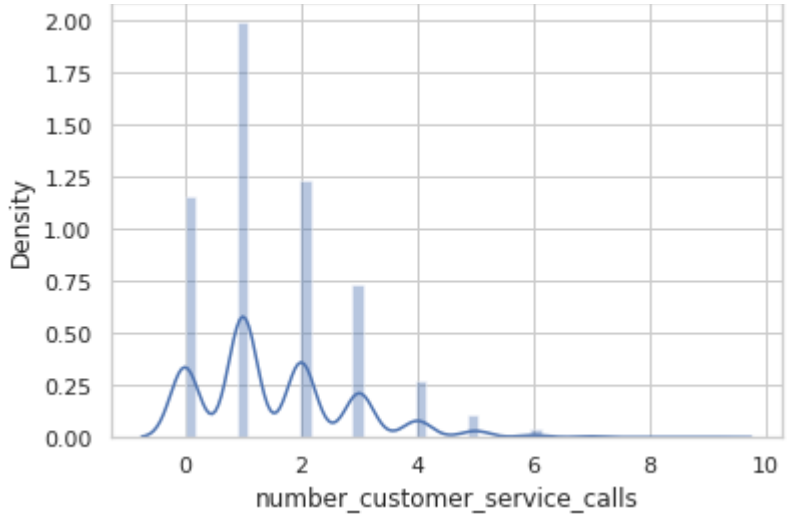

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```




```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



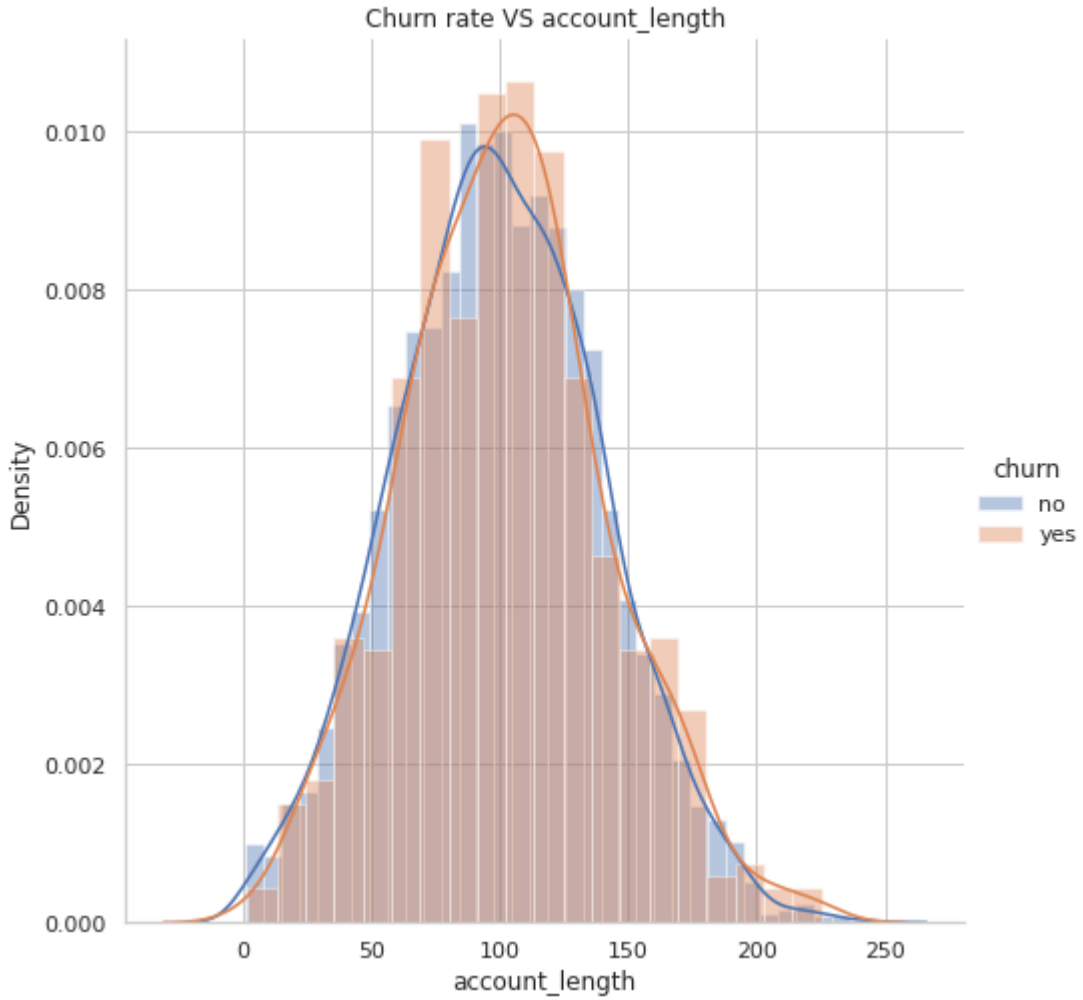

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```


2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Density
0
2
4
6
8
10
number_customer_service_calls

BiVariate Analysis

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'account_length').add_legend()
plt.title('Churn rate VS account_length')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
Text(0.5, 1.0, 'Churn rate VS account_length')
```

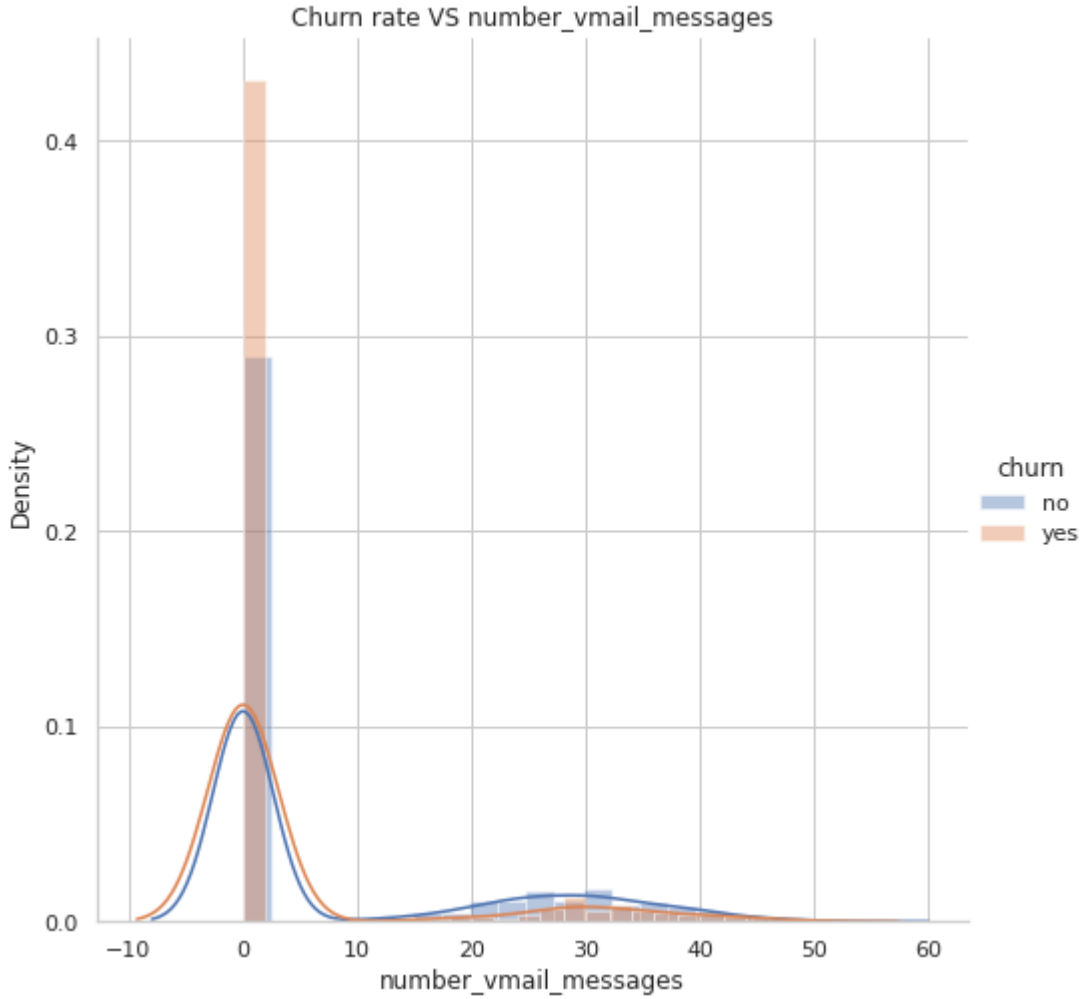


customers account length between 60 to 120 has more churn rate

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'number_vmail_messages').add_legend()
```

```
plt.title('Churn rate VS number_vmail_messages')
plt.show()
```

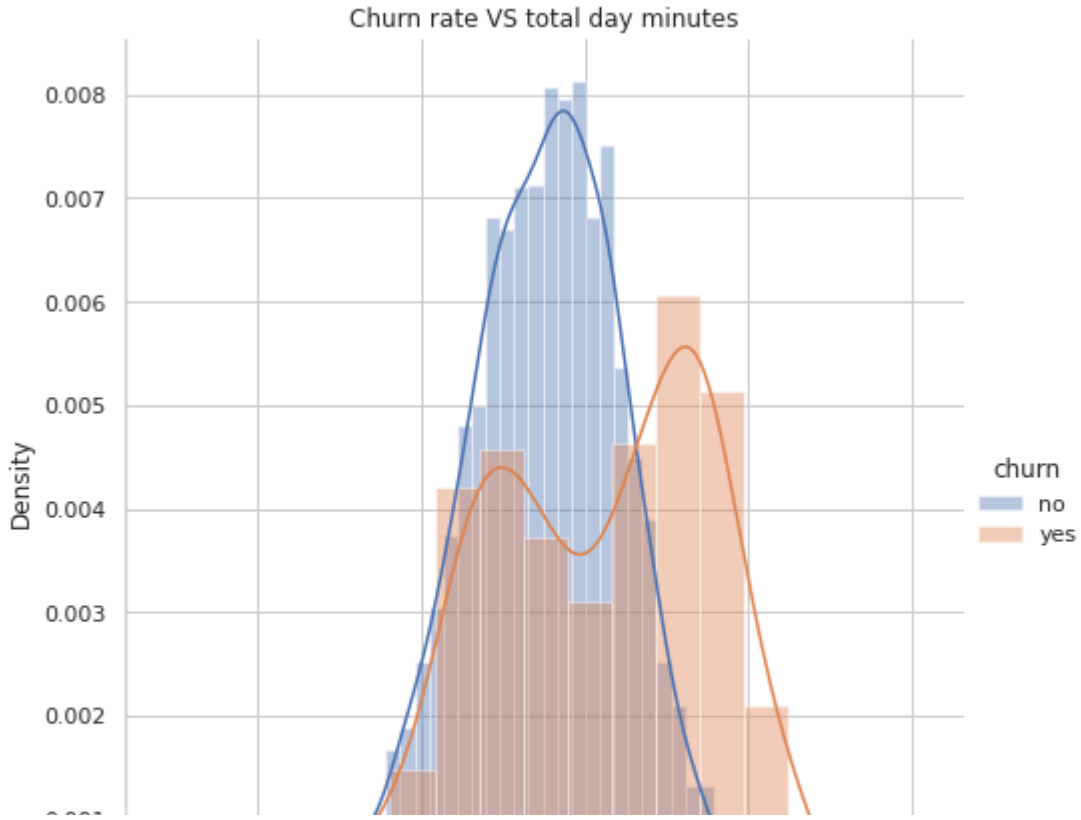
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



More churn rate when the number_vamil_messages is 0

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_day_minutes').add_legend()
plt.title('Churn rate VS total day minutes')
plt.show()
```

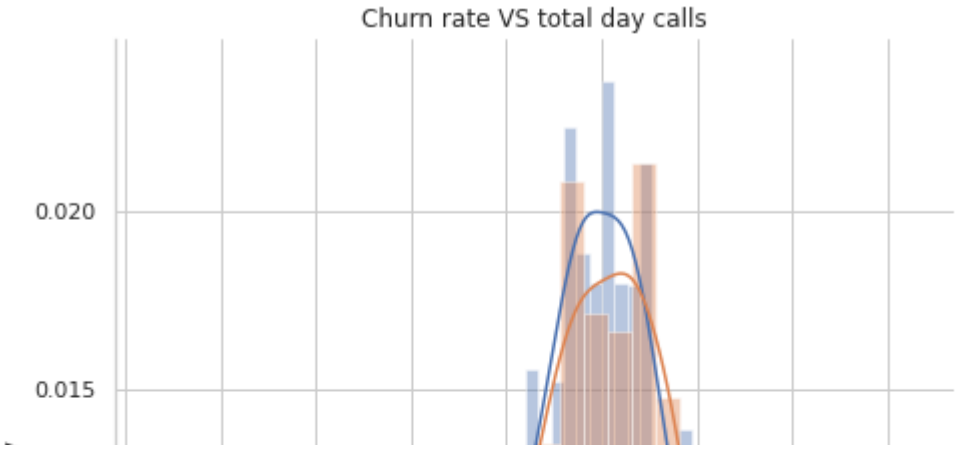
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



Churn rate is high when the total_day_minutes is lies between 210 min to 300 min.

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_day_calls').add_legend()
plt.title('Churn rate VS total day calls')
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



churn rate is high lies between 85 to 115.

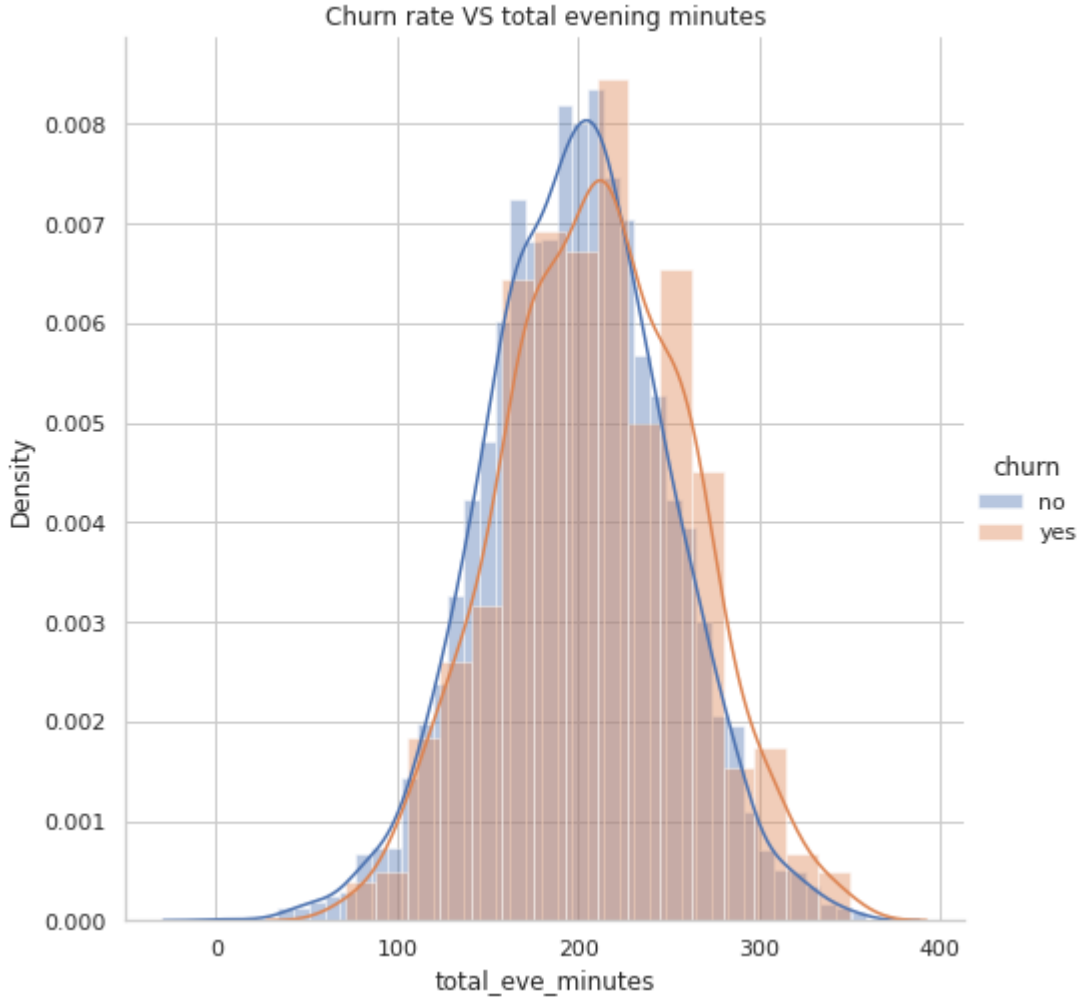
```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_day_charge').add_legend()
plt.title('Churn rate VS total day charge')
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```

churn rate is high when total day charge is lies between 40 to 50.

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_eve_minutes').add_legend()
plt.title('Churn rate VS total evening minutes')
plt.show()
```

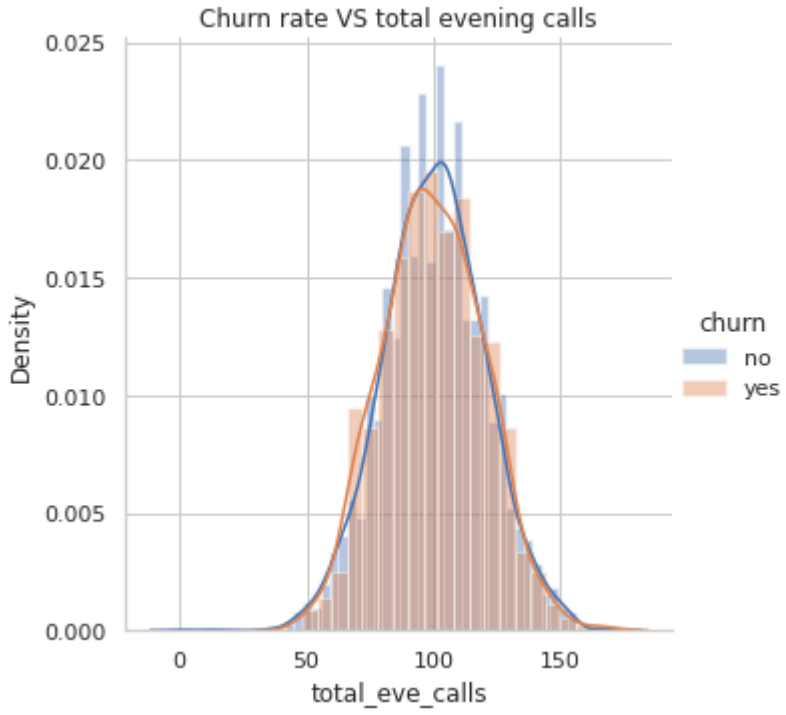
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



churn rate is high when the total evening minutes is lies between 180 min to 220 min

```
sns.FacetGrid(df, hue='churn',size=5).map(sns.distplot, 'total_eve_calls').add_legend()
plt.title('Churn rate VS total evening calls')
plt.show()
```

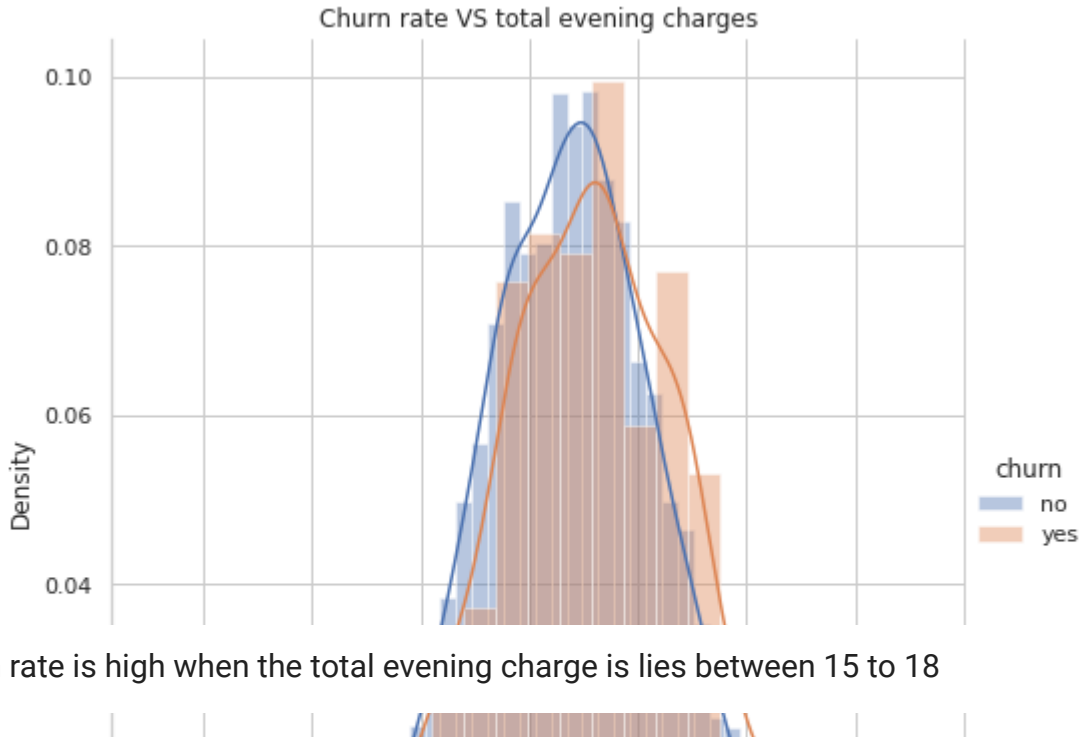
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



churn rate is high when total evening calls lies between 90 to 115.

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_eve_charge').add_legend()
plt.title('Churn rate VS total evening charges')
plt.show()
```

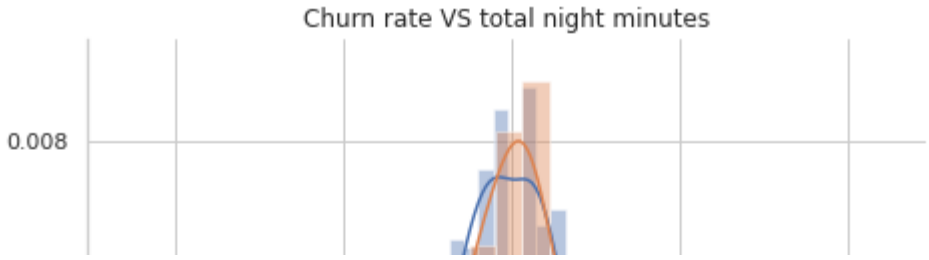
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



churn rate is high when the total evening charge is lies between 15 to 18

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_night_minutes').add_legend()
plt.title('Churn rate VS total night minutes')
plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337: UserWarning: The `size`
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `di
  warnings.warn(msg, FutureWarning)
```



Churn rate is high when the total_night_minutes is lies between 190 to 220 min

```
sns.FacetGrid(df, hue='churn',size=7).map(sns.distplot, 'total_night_calls').add_legend()
plt.title('Churn rate VS total night calls')
plt.show()
```