In [1]:
```
import pandas as pd
import numpy as np
import plotly.express as px
import plotly.graph_objects as go
```

In [2]:
```
data = pd.read_csv("D:/tips.csv")
```

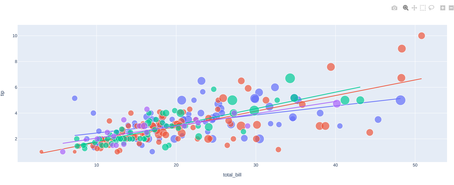
In [3]:
```
print(data.head())
```

```
   total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
```

In [4]:
```
data.isnull().sum()
```

Out[4]:
```
total_bill    0
tip           0
sex           0
smoker        0
day           0
time          0
size          0
dtype: int64
```

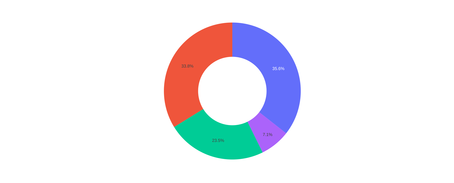
In [5]:
```
figure = px.scatter(data_frame = data, x="total_bill",
                    y="tip", size="size", color= "day", trendline="ols")
figure.show()
```



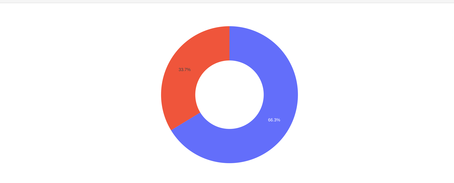
In [6]:
```
figure = px.scatter(data_frame = data, x="total_bill",
                    y="tip", size="size", color= "sex", trendline="ols")
figure.show()
```

In [7]:
```
figure = px.scatter(data_frame = data, x="total_bill",
                    y="tip", size="size", color= "time", trendline="ols")
figure.show()
```

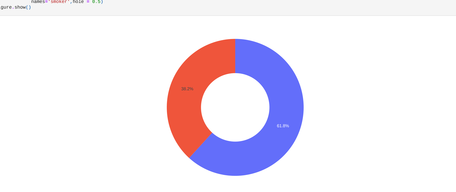
In [8]:
```
figure = px.pie(data,
                values="tip",
                names='day',hole = 0.5)
figure.show()
```



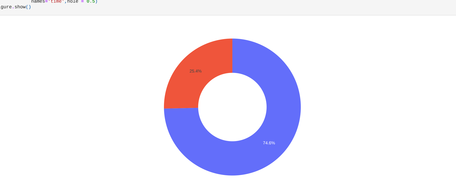
In [9]:
```
figure = px.pie(data,
                values="tip",
                names="sex",hole = 0.5)
figure.show()
```



In [10]:
```
figure = px.pie(data,
                values="tip",
                names='smoker',hole = 0.5)
figure.show()
```



In [11]:
```
figure = px.pie(data,
                values="tip",
                names='time',hole = 0.5)
figure.show()
```



In [12]:
```
data["sex"] = data["sex"].map({"Female": 0, "Male": 1})
data["smoker"] = data["smoker"].map({"No": 0, "Yes": 1})
data['day'] = data['day'].map({'Thur': 0, 'Fri': 1, 'Sat': 2, 'Sun': 3})
data["time"] = data["time"].map({"Lunch": 0, "Dinner": 1})
```

In [13]:
```
data.head()
```

Out[13]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	0	0	3	1	2
1	10.34	1.66	1	0	3	1	3
2	21.01	3.50	1	0	3	1	3
3	23.68	3.31	1	0	3	1	2
4	24.59	3.61	0	0	3	1	4

In [14]:
```
x = np.array(data[["total_bill", "sex", "smoker", "day",
                   "time", "size"]])
y = np.array(data["tip"])
```

In [15]:
```
from sklearn.model_selection import train_test_split
```

In [16]:
```
xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.3,
                                                random_state=42)
```

In [17]:
```
from sklearn.linear_model import LinearRegression
```

In [18]:
```
from sklearn.linear_model import LogisticRegression
```

In [19]:
```
x = np.array(data[["total_bill", "sex", "smoker", "day",
                   "time", "size"]])
y = np.array(data["tip"])

from sklearn.model_selection import train_test_split
xtrain, xtest, ytrain, ytest = train_test_split(x, y,
                                                test_size=0.2,
                                                random_state=42)
```

In [20]:
```
model = LinearRegression()
```

In [21]:
```
model.fit(xtrain, ytrain)
```

Out[21]:
```
LinearRegression()
```

In [22]:
```
# features = [[total_bill, "sex", "smoker", "day", "time", "size"]]
features = np.array([[24.50, 1, 0, 0, 1, 4]])
model.predict(features)
```

Out[22]:
```
array([3.73742609])
```

In [23]:
```
y_pred = model.predict(xtest)
```

In [24]:
```
model.score(xtest,ytest)
```

Out[24]:
```
0.4429399687489898
```

In [25]:
```
import math
from sklearn.metrics import mean_squared_error
```

In [26]:
```
rmse = math.sqrt(mean_squared_error(ytest,y_pred))
```

In [27]:
```
print(rmse)
```

```
0.8344513626960923
```

In [28]:
```
from sklearn.metrics import mean_absolute_error
```

In [29]:
```
mean_absolute_error(ytest, y_pred)
```

Out[29]:
```
0.6685728166722872
```

In [30]:
```
from sklearn.metrics import r2_score
r2_score(ytest,y_pred)
```

Out[30]:
```
0.4429399687489898
```

In [ ]: