In [1]:
```
import pandas as pd
import glob
import os
import matplotlib.pyplot as plt
joined_files = os.path.join(r"C:\Users\dvanama\Documents\Uber-dataset\Uber-datase
                            "uber-raw-data*.csv")
joined_list = glob.glob(joined_files)
df = pd.concat(map(pd.read_csv, joined_list), ignore_index=True)
df
```

Out[1]:

	Date/Time	Lat	Lon	Base
0	4/1/2014 0:11:00	40.7690	-73.9549	B02512
1	4/1/2014 0:17:00	40.7267	-74.0345	B02512
2	4/1/2014 0:21:00	40.7316	-73.9873	B02512
3	4/1/2014 0:28:00	40.7588	-73.9776	B02512
4	4/1/2014 0:33:00	40.7594	-73.9722	B02512
...	...	...	...	...
4534322	9/30/2014 22:57:00	40.7668	-73.9845	B02764
4534323	9/30/2014 22:57:00	40.6911	-74.1773	B02764
4534324	9/30/2014 22:58:00	40.8519	-73.9319	B02764
4534325	9/30/2014 22:58:00	40.7081	-74.0066	B02764
4534326	9/30/2014 22:58:00	40.7140	-73.9496	B02764

4534327 rows × 4 columns

In [13]:
```
df['Date/Time'] = pd.to_datetime(df['Date/Time'])
df['Month'] = df['Date/Time'].dt.month
df['HourOfDay'] = df['Date/Time'].dt.hour
q2=df.groupby(by='Month').count()
q3=df.groupby(by='Base').count()
q2
```

Out[13]:

Month	Date/Time	Lat	Lon	Base	HourOfDay
4	564516	564516	564516	564516	564516
5	652435	652435	652435	652435	652435
6	663844	663844	663844	663844	663844
7	796121	796121	796121	796121	796121
8	829275	829275	829275	829275	829275
9	1028136	1028136	1028136	1028136	1028136

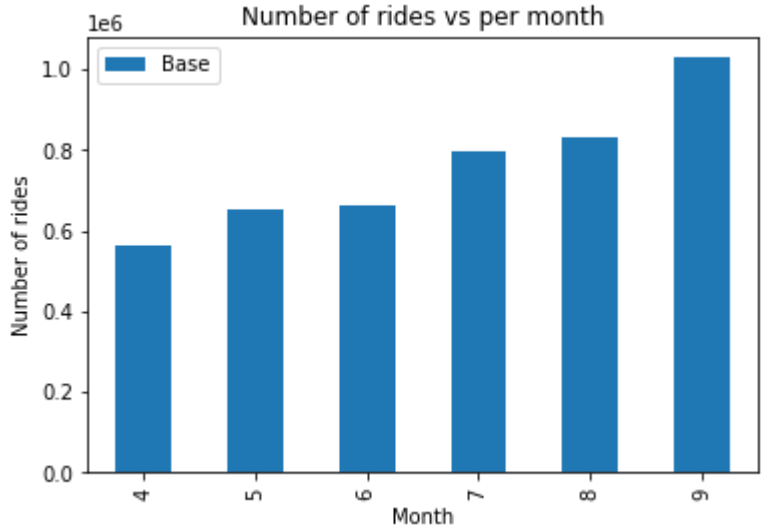
In [17]:
```
trips =  df.groupby('Base')['Date/Time'].count()
q=pd.DataFrame(trips)
q
```

Out[17]:

Base	Date/Time
B02512	205673
B02598	1393113
B02617	1458853
B02682	1212789
B02764	263899

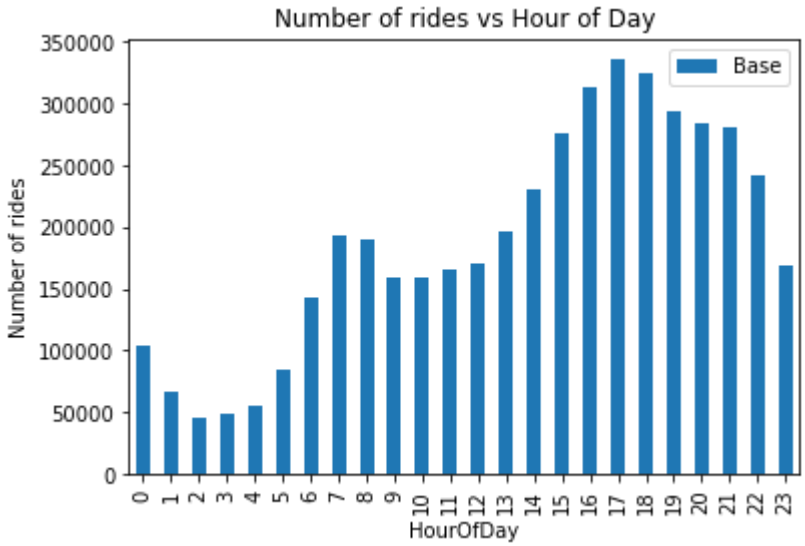
In [14]:
```
Months=df.pivot_table(index=['Month'],
                      values="Base",
                      aggfunc="count")
Months.plot(kind="bar")
plt.ylabel("Number of rides")
plt.title("Number of rides vs per month")
```

Out[14]:
```
Text(0.5, 1.0, 'Number of rides vs per month')
```



In [15]:
```
Hours=df.pivot_table(index=["HourOfDay"],values="Base",aggfunc="count")
Hours.plot(kind='bar')
plt.ylabel("Number of rides")
plt.title("Number of rides vs Hour of Day")
```

Out[15]: Text(0.5, 1.0, 'Number of rides vs Hour of Day')



In [19]:
```
pip install pandasql
```

```
Requirement already satisfied: pandasql in c:\users\dvanama\anaconda3\lib\site-
packages (0.7.3)
Requirement already satisfied: numpy in c:\users\dvanama\anaconda3\lib\site-pac
kages (from pandasql) (1.20.1)
Requirement already satisfied: pandas in c:\users\dvanama\anaconda3\lib\site-pa
ckages (from pandasql) (1.2.4)
Requirement already satisfied: sqlalchemy in c:\users\dvanama\anaconda3\lib\sit
e-packages (from pandasql) (1.4.7)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\dvanama\anaco
nda3\lib\site-packages (from pandas->pandasql) (2.8.1)
Requirement already satisfied: pytz>=2017.3 in c:\users\dvanama\anaconda3\lib\s
ite-packages (from pandas->pandasql) (2021.1)
Requirement already satisfied: six>=1.5 in c:\users\dvanama\anaconda3\lib\site-
packages (from python-dateutil>=2.7.3->pandas->pandasql) (1.15.0)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\dvanama\anaconda3\l
ib\site-packages (from sqlalchemy->pandasql) (1.0.0)
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
import pandas as pd
from pandasql import sqldf
pysqldf = lambda q: sqldf(q,globals())
```

In [4]:
```
def open_file(x,y =','):
    de= pd.read_csv(x,index_col=0,sep=y)
    return de
```

In [5]:
```
loan = open_file('C:\dr\Loan_Prediction_Dataset.csv')
loan
```

Out[5]:

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIı
LP001015	Male	Yes	0	Graduate	No	5720	
LP001022	Male	Yes	1	Graduate	No	3076	
LP001031	Male	Yes	2	Graduate	No	5000	
LP001035	Male	Yes	2	Graduate	No	2340	
LP001051	Male	No	0	Not Graduate	No	3276	
...	...	...	...	...	...	...	
LP002971	Male	Yes	3+	Not Graduate	Yes	4009	
LP002975	Male	Yes	0	Graduate	No	4158	
LP002980	Male	No	0	Graduate	No	3250	
LP002986	Male	Yes	0	Graduate	No	5000	
LP002989	Male	No	0	Graduate	Yes	9200	

367 rows × 11 columns

In [6]:
```
q= loan.isna().mean().round(4)*100
q
```

Out[6]:
```
Gender               3.00
Married              0.00
Dependents           2.72
Education            0.00
Self_Employed        6.27
ApplicantIncome      0.00
CoapplicantIncome    0.00
LoanAmount           1.36
Loan_Amount_Term     1.63
Credit_History       7.90
Property_Area        0.00
dtype: float64
```

In [23]:
```
q2 = sqldf("select Gender,count(self_Employed) from loan where Gender ='Female' A
q2
```

Out[23]:

	Gender	count(self_Employed)
0	Female	4

In [31]:
```
q3=sqldf("select Gender,count(Married),Property_Area from loan where Gender ='Mal
```

In [32]:
```
q3
```

Out[32]:

	Gender	count(Married)	Property_Area
0	Male	26	Semiurban

In [33]:
```
q4= sqldf("select Max(CoapplicantIncome) from loan where Education = 'Graduate'")
q4
```

Out[33]:

	Max(CoapplicantIncome)
0	24000

In [34]:
```
q5 = sqldf("select AVG(LoanAmount) from loan where Dependents ='3+'")
q5
```

Out[34]:

	AVG(LoanAmount)
0	146.35

In [ ]:
```
```