In [43]:
```
import pandas as pd
import numpy as np
from matplotlib import pyplot
from tensorflow import keras
import matplotlib.pyplot as plt
```

UK'de 2000-2016 yıllarına ait 1.6 milyondan fazla trafik kazası

1. Veri önişleme

In [5]:
```
# veri yükleme işlemi
acc_05_07 = pd.read_csv("accidents_2005_to_2007.csv", index_col='Accident_Index', pa
acc_09_11 = pd.read_csv("accidents_2009_to_2011.csv", index_col='Accident_Index', pa
acc_12_14 = pd.read_csv("accidents_2012_to_2014.csv", index_col='Accident_Index', pa
```

In [8]:
```
acc_12_14[0:10]
```

Out[8]:

Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude	Latitude	Police_Force
201201BS70001	527200	178760	-0.169101	51.493429	1
201201BS70002	524930	181430	-0.200838	51.517931	1
201201BS70003	525860	178080	-0.188636	51.487618	1
201201BS70004	524980	181030	-0.200259	51.514325	1
201201BS70005	526170	179200	-0.183773	51.497614	1
201201BS70006	526090	177600	-0.185496	51.483253	1
201201BS70007	527780	179680	-0.160418	51.501567	1
201201BS70008	524010	182080	-0.213862	51.523975	1
201201BS70010	527710	179290	-0.161567	51.498077	1
201201BS70011	525120	180060	-0.198587	51.505576	1

10 rows × 32 columns

In [10]:
```
# 3 veri setini alt alta birleştirme
acc_05_14 = pd.concat([acc_05_07, acc_09_11, acc_12_14])
```

In [11]:
```
acc_05_14.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1504150 entries, 200501BS00001 to 2.01E+12
Data columns (total 32 columns):
 #   Column                                       Non-Null Count    Dtype  
---  ------                                       --------------    -----  
 0   Location_Easting_OSGR                        1504049 non-null  float64
 1   Location_Northing_OSGR                       1504049 non-null  float64
 2   Longitude                                    1504049 non-null  float64
 3   Latitude                                     1504049 non-null  float64
 4   Police_Force                                 1504150 non-null  int64  
 5   Accident_Severity                            1504150 non-null  int64  
 6   Number_of_Vehicles                           1504150 non-null  int64  
 7   Number_of_Casualties                         1504150 non-null  int64  
 8   Date                                         1504150 non-null  object 
 9   Day_of_Week                                  1504150 non-null  int64  
 10  Time                                         1504033 non-null  object 
 11  Local_Authority_(District)                   1504150 non-null  int64  
 12  Local_Authority_(Highway)                    1504150 non-null  object 
 13  1st_Road_Class                               1504150 non-null  int64  
 14  1st_Road_Number                              1504150 non-null  int64  
 15  Road_Type                                    1504150 non-null  object 
 16  Speed_limit                                  1504150 non-null  int64  
 17  Junction_Detail                              0 non-null        float64
 18  Junction_Control                             901315 non-null   object 
 19  2nd_Road_Class                               1504150 non-null  int64  
 20  2nd_Road_Number                              1504150 non-null  int64  
 21  Pedestrian_Crossing-Human_Control            1504133 non-null  object 
 22  Pedestrian_Crossing-Physical_Facilities      1504116 non-null  object 
 23  Light_Conditions                             1504150 non-null  object 
 24  Weather_Conditions                           1504024 non-null  object 
 25  Road_Surface_Conditions                      1502192 non-null  object 
 26  Special_Conditions_at_Site                   1504135 non-null  object 
 27  Carriageway_Hazards                          1504121 non-null  object 
 28  Urban_or_Rural_Area                          1504150 non-null  int64  
 29  Did_Police_Officer_Attend_Scene_of_Accident  1501228 non-null  object 
 30  LSOA_of_Accident_Location                    1395912 non-null  object 
 31  Year                                         1504150 non-null  int64  
dtypes: float64(5), int64(13), object(14)
memory usage: 378.7+ MB
```

In [13]:
```
# date ve time kolanları corelation.

acc_05_14['date_time'] = acc_05_14['Date']+ ' ' + acc_05_14['Time']
```

In [15]:
```
time_format = '%d/%m/%Y %H:%M'
acc_05_14['date_time'] = pd.to_datetime(acc_05_14['date_time'], format=time_format)
```

In [17]:
```
kopya_satir = acc_05_14.duplicated()      # kopya varsa true dönecektir...
print(kopya_satir.any())
```

```
# df.drop_duplicates(inplace=True)   benzer satırları silmek için kullanılır
```

```
True
```

In [18]:
```
acc_05_14.drop_duplicates(inplace=True)
```

In [20]:
```
print(acc_05_14.nunique()) # sutunlarda yeralan verilerin benzersiz sayıları
```

```
Location_Easting_OSGR                          182519
Location_Northing_OSGR                         221877
Longitude                                     1059046
Latitude                                      1001148
Police_Force                                       51
Accident_Severity                                   3
Number_of_Vehicles                                 27
Number_of_Casualties                               47
Date                                             3286
Day_of_Week                                         7
Time                                             1439
Local_Authority_(District)                        416
Local_Authority_(Highway)                         207
1st_Road_Class                                      6
1st_Road_Number                                  6854
Road_Type                                           6
Speed_limit                                         8
Junction_Detail                                     0
Junction_Control                                    4
2nd_Road_Class                                      7
2nd_Road_Number                                  7235
Pedestrian_Crossing-Human_Control                   3
Pedestrian_Crossing-Physical_Facilities             6
Light_Conditions                                    5
Weather_Conditions                                  9
Road_Surface_Conditions                             5
Special_Conditions_at_Site                          8
Carriageway_Hazards                                 6
Urban_or_Rural_Area                                 3
Did_Police_Officer_Attend_Scene_of_Accident         2
LSOA_of_Accident_Location                       35452
Year                                                9
date_time                                      914222
dtype: int64
```

In [27]:
```
print(acc_05_14.describe())  # dataset ile ilgili istatiksel bilgiler görüntülenir.
```

```
       Location_Easting_OSGR  Location_Northing_OSGR     Longitude  \
count           1.469882e+06            1.469882e+06  1.469882e+06
mean            4.398981e+05            2.986733e+05 -1.432638e+00
std             9.553458e+04            1.612596e+05  1.404319e+00
min             6.495000e+04            1.029000e+04 -7.516225e+00
25%             3.757500e+05            1.780065e+05 -2.363659e+00
50%             4.409300e+05            2.653400e+05 -1.391630e+00
75%             5.232900e+05            3.966000e+05 -2.184805e-01
max             6.553700e+05            1.208800e+06  1.759398e+00

           Latitude  Police_Force  Accident_Severity  Number_of_Vehicles  \
count  1.469882e+06  1.469983e+06       1.469983e+06        1.469983e+06
mean   5.257595e+01  3.078162e+01       2.838772e+00        1.831849e+00
std    1.452062e+00  2.551800e+01       4.014223e-01        7.152255e-01
min    4.991294e+01  1.000000e+00       1.000000e+00        1.000000e+00
```

```
25%    5.148790e+01  7.000000e+00       3.000000e+00        1.000000e+00
50%    5.227670e+01  3.100000e+01       3.000000e+00        2.000000e+00
75%    5.346435e+01  4.600000e+01       3.000000e+00        2.000000e+00
max    6.075754e+01  9.800000e+01       3.000000e+00        6.700000e+01

       Number_of_Casualties   Day_of_Week  Local_Authority_(District)  \
count          1.469983e+06  1.469983e+06                1.469983e+06
mean           1.350894e+00  4.118635e+00                3.535679e+02
std            8.258026e-01  1.924706e+00                2.592777e+02
min            1.000000e+00  1.000000e+00                1.000000e+00
25%            1.000000e+00  2.000000e+00                1.220000e+02
50%            1.000000e+00  4.000000e+00                3.280000e+02
75%            1.000000e+00  6.000000e+00                5.320000e+02
max            9.300000e+01  7.000000e+00                9.410000e+02

       1st_Road_Class  1st_Road_Number   Speed_limit  Junction_Detail  \
count    1.469983e+06     1.469983e+06  1.469983e+06              0.0
mean     4.089807e+00     1.008857e+03  3.907890e+01              NaN
std      1.429899e+00     1.821696e+03  1.417132e+01              NaN
min      1.000000e+00    -1.000000e+00  1.000000e+01              NaN
25%      3.000000e+00     0.000000e+00  3.000000e+01              NaN
50%      4.000000e+00     1.290000e+02  3.000000e+01              NaN
75%      6.000000e+00     7.260000e+02  5.000000e+01              NaN
max      6.000000e+00     9.999000e+03  7.000000e+01              NaN

       2nd_Road_Class  2nd_Road_Number  Urban_or_Rural_Area          Year
count    1.469983e+06     1.469983e+06         1.469983e+06  1.469983e+06
mean     2.663929e+00     3.800842e+02         1.356720e+00  2.009309e+03
std      3.207805e+00     1.300904e+03         4.792189e-01  3.021204e+00
min     -1.000000e+00    -1.000000e+00         1.000000e+00  2.005000e+03
25%     -1.000000e+00     0.000000e+00         1.000000e+00  2.006000e+03
50%      3.000000e+00     0.000000e+00         1.000000e+00  2.010000e+03
75%      6.000000e+00     0.000000e+00         2.000000e+00  2.012000e+03
max      6.000000e+00     9.999000e+03         3.000000e+00  2.014000e+03
```

In [26]:

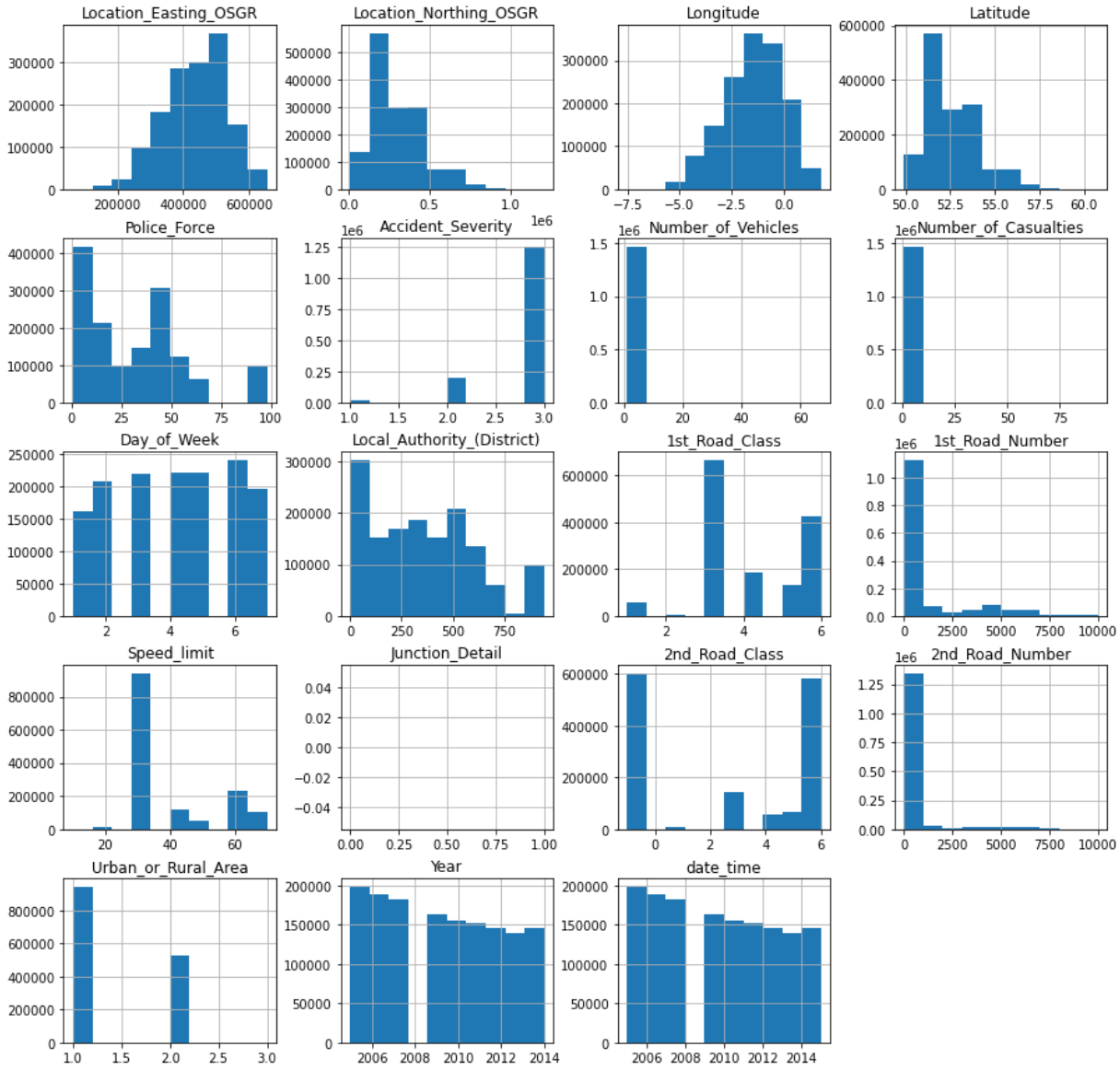
Out[26]:

	Location_Easting_OSGR	Location_Northing_OSGR	Longitude	Latitude	Police_Force
count	1.469882e+06	1.469882e+06	1.469882e+06	1.469882e+06	1.469983e+06
mean	4.398981e+05	2.986733e+05	1.432638e+00	5.257595e+01	3.078162e+01
std	9.553458e+04	1.612596e+05	1.404319e+00	1.452062e+00	2.551800e+01
min	6.495000e+04	1.029000e+04	7.516225e+00	4.991294e+01	1.000000e+00
25%	3.757500e+05	1.780065e+05	2.363659e+00	5.148790e+01	7.000000e+00
50%	4.409300e+05	2.653400e+05	1.391630e+00	5.227670e+01	3.100000e+01
75%	5.232900e+05	3.966000e+05	2.184805e-01	5.346435e+01	4.600000e+01
max	6.553700e+05	1.208800e+06	1.759398e+00	6.075754e+01	9.800000e+01

In [28]:

```
acc_05_14.hist(figsize=(15,15))
pyplot.show()

# eğer çok fazla kolon normal dağılıma uymuyorsa transform işlemi yapılabilir.
```



In [29]:
```
acc_05_14.shape
```

Out[29]:
```
(1469983, 33)
```

In [30]:
```
print(acc_05_14.isnull().sum())
```

```
Location_Easting_OSGR                101
Location_Northing_OSGR               101
Longitude                            101
Latitude                             101
Police_Force                           0
Accident_Severity                      0
Number_of_Vehicles                     0
Number_of_Casualties                   0
Date                                   0
Day_of_Week                            0
Time                                 117
Local_Authority_(District)             0
Local_Authority_(Highway)              0
1st_Road_Class                         0
1st_Road_Number                        0
Road_Type                              0
Speed_limit                            0
Junction_Detail                  1469983
Junction_Control                  591811
```

```
2nd_Road_Class                                  0
2nd_Road_Number                                 0
Pedestrian_Crossing-Human_Control              17
Pedestrian_Crossing-Physical_Facilities        34
Light_Conditions                                0
Weather_Conditions                            126
Road_Surface_Conditions                      1945
Special_Conditions_at_Site                     15
Carriageway_Hazards                            29
Urban_or_Rural_Area                             0
Did_Police_Officer_Attend_Scene_of_Accident    2922
LSOA_of_Accident_Location                    108229
Year                                            0
date_time                                     117
dtype: int64
```

In [32]:
```
# Junction_Detail, ve  Junction_Control kolonları tamamen boşlar.
acc_05_14 = acc_05_14.drop(columns=["Junction_Control","Junction_Detail"])
```

In [33]:
```
acc_05_14.shape
```

Out[33]:
```
(1469983, 31)
```

In [39]:
```
# boş olan hucreler satırı az o yuzden atalım (e.g. langıtite 101)
acc_05_14.dropna(subset=['Location_Easting_OSGR'], inplace=True)
```

In [40]:
```
print(acc_05_14.isnull().sum())
```

```
Location_Easting_OSGR                          0
Location_Northing_OSGR                         0
Longitude                                      0
Latitude                                       0
Police_Force                                   0
Accident_Severity                              0
Number_of_Vehicles                             0
Number_of_Casualties                           0
Date                                           0
Day_of_Week                                    0
Time                                         117
Local_Authority_(District)                     0
Local_Authority_(Highway)                      0
1st_Road_Class                                 0
1st_Road_Number                                0
Road_Type                                      0
Speed_limit                                    0
2nd_Road_Class                                 0
2nd_Road_Number                                0
Pedestrian_Crossing-Human_Control             17
Pedestrian_Crossing-Physical_Facilities       33
Light_Conditions                               0
Weather_Conditions                           126
Road_Surface_Conditions                     1944
Special_Conditions_at_Site                    15
Carriageway_Hazards                           29
Urban_or_Rural_Area                            0
Did_Police_Officer_Attend_Scene_of_Accident    2921
LSOA_of_Accident_Location                    108128
Year                                           0
```

```
date_time                                   117
dtype: int64
```

In [41]:

```
acc_05_14.head()
```

Out[41]:

Accident_Index	Location_Easting_OSGR	Location_Northing_OSGR	Longitude	Latitude	Police_Force
200501BS00001	525680.0	178240.0	-0.191170	51.489096	1
200501BS00002	524170.0	181650.0	-0.211708	51.520075	1
200501BS00003	524520.0	182240.0	-0.206458	51.525301	1
200501BS00004	526900.0	177530.0	-0.173862	51.482442	1
200501BS00005	528060.0	179040.0	-0.156618	51.495752	1

5 rows × 31 columns

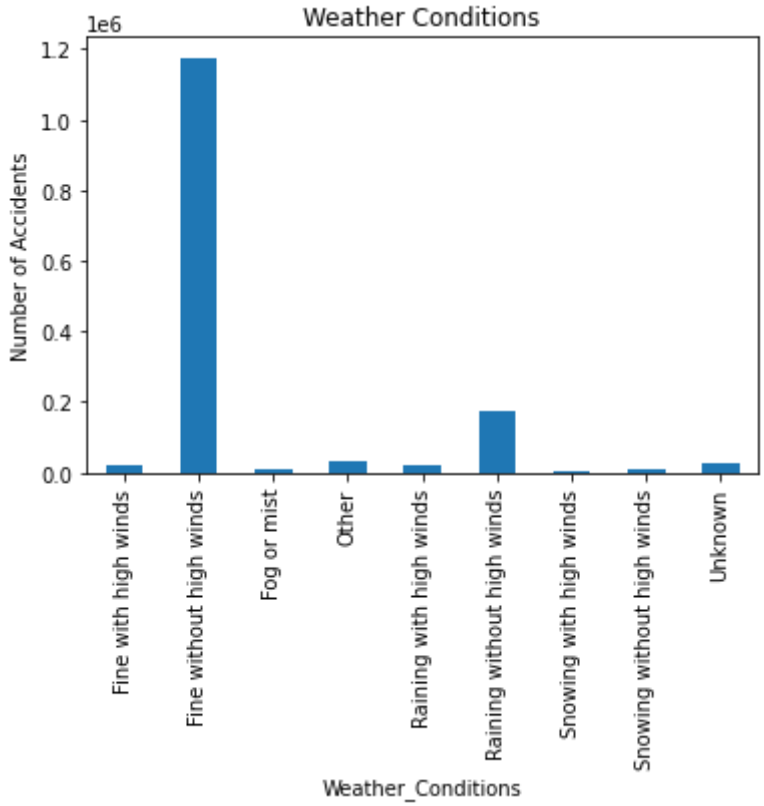
In [49]:

```
# hava koşullarına göre kaza sayıları

acc_05_14['Weather_Conditions'].unique()
acc_05_14.groupby('Weather_Conditions').size().plot(kind = 'bar')
plt.title('Weather Conditions')
plt.ylabel('Number of Accidents')
```

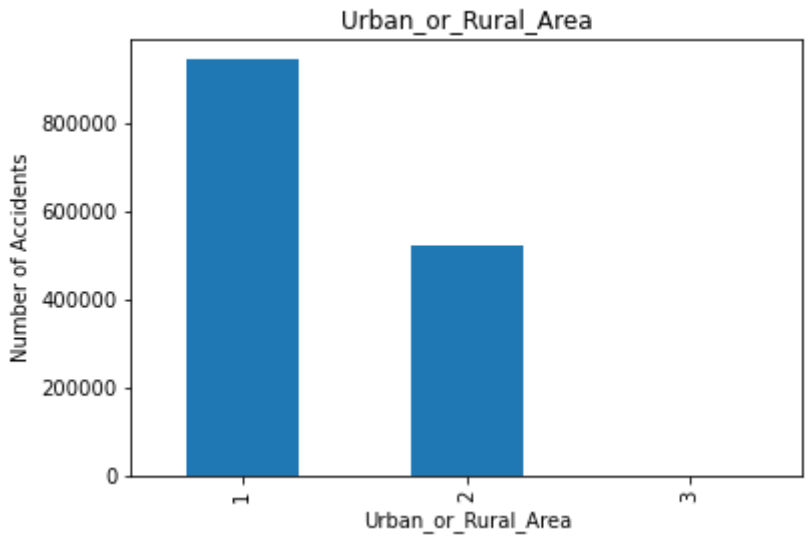
Out[49]:

```
Text(0, 0.5, 'Number of Accidents')
```



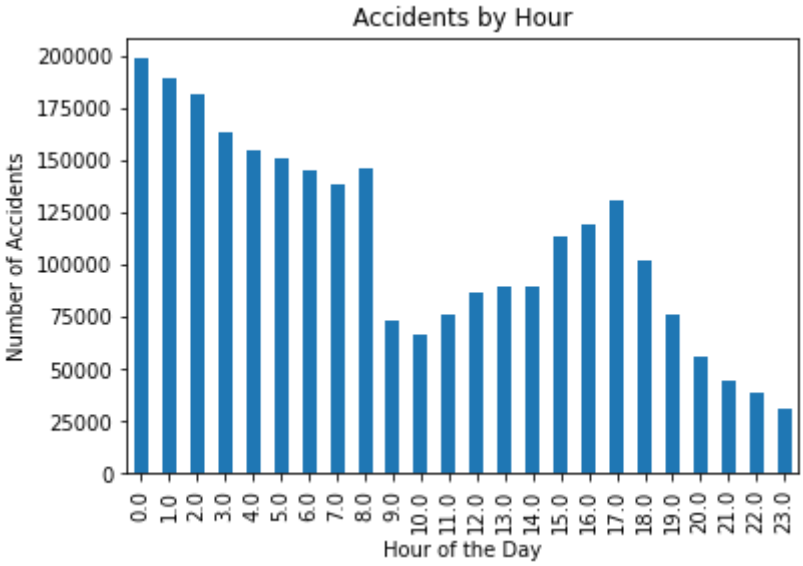
In [50]:
```
acc_05_14['Urban_or_Rural_Area'].unique()
acc_05_14.groupby('Urban_or_Rural_Area').size().plot(kind = 'bar')
plt.title('Urban_or_Rural_Area')
plt.ylabel('Number of Accidents')
```

Out[50]:
```
Text(0, 0.5, 'Number of Accidents')
```



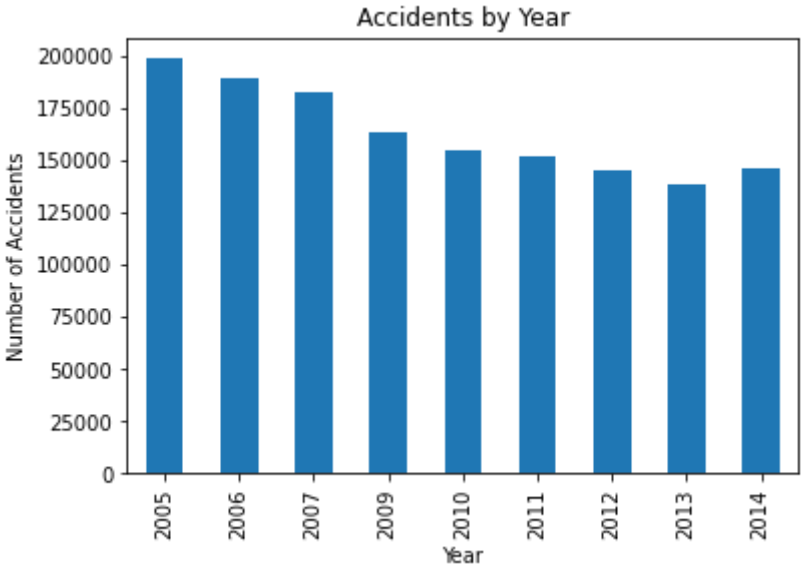
In [46]:
```
#Günlük kaza sayıları
acc_05_14.groupby(acc_05_14['date_time'].dt.hour).size().plot(kind='bar', title='Acc
plt.ylabel('Number of Accidents')
plt.xlabel('Hour of the Day')
```

Out[46]:
```
Text(0.5, 0, 'Hour of the Day')
```



In [47]:
```
# yıllık kaza sayıları
acc_05_14.groupby('Year').size().plot(kind='bar', title='Accidents by Year')
plt.ylabel('Number of Accidents')
```

Out[47]:
```
Text(0, 0.5, 'Number of Accidents')
```



In [ ]: