In [2]:

```
#to calc most world cup winning team
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var=pd.read_csv('WorldCups.csv')
var=var.fillna('unavailable')
a=var['Winner'].value_counts()
a=(pd.DataFrame({'name':a.index, 'times_won':a.values})).head(1)
a
```

Out[2]:

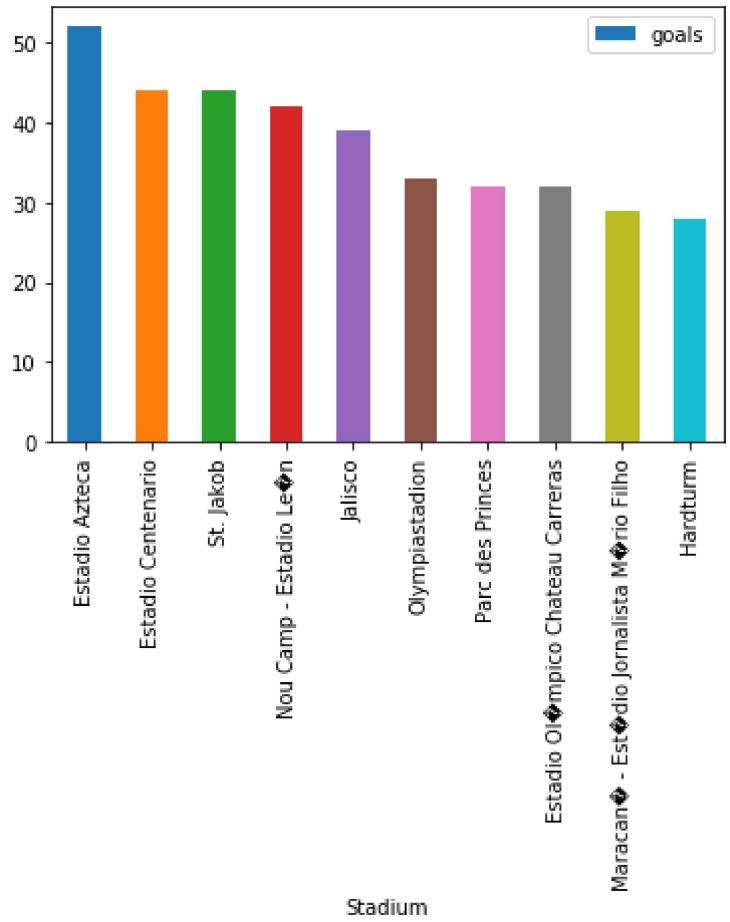
	name	times_won
0	Brazil	5

In [4]:

```
#top 10 stadium based on goals
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var1=pd.read_csv('WorldCupMatches.csv')
var1=var1.fillna(0)
var1['totalgoals']=var1['Home Team Goals']+var1['Away Team Goals']
d=var1.groupby('Stadium')['totalgoals'].sum()
d=d.sort_values(ascending=False)
d=d.head(10)
d=pd.DataFrame({'Stadium':d.index, 'goals':d.values})
plot1 = d.plot.bar(x='Stadium',y='goals', figsize=(6,4))
plot1
```

Out[4]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218be6f09b0>
```

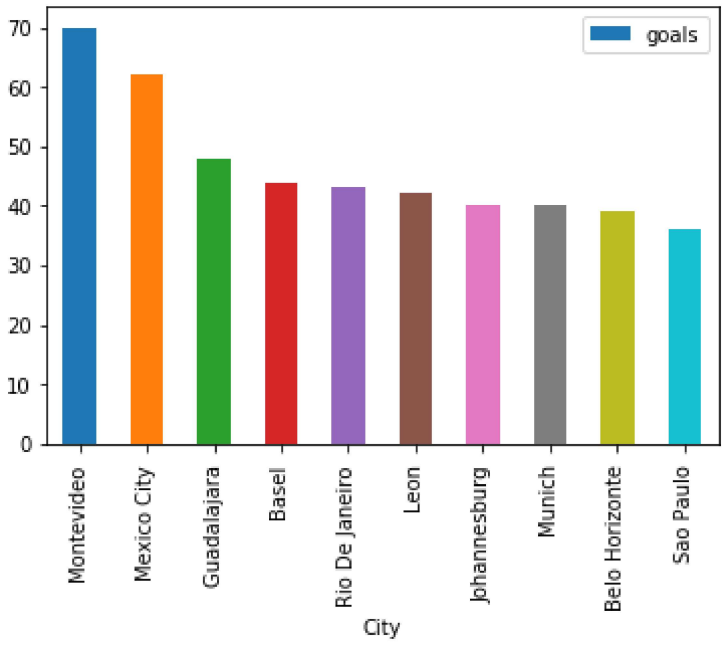


In [9]:

```
#top 10 city based on goals
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var2=pd.read_csv('WorldCupMatches.csv')
var2=var2.fillna(0)
var2['totalgoals']=var2['Home Team Goals']+var2['Away Team Goals']
var2
e=var2.groupby('City')['totalgoals'].sum()
e=e.sort_values(ascending=False)
e=e.head(10)
var2
e=pd.DataFrame({'City':e.index, 'goals':e.values})
e.plot.bar(x='City',y='goals')
```

Out[9]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218be520208>
```

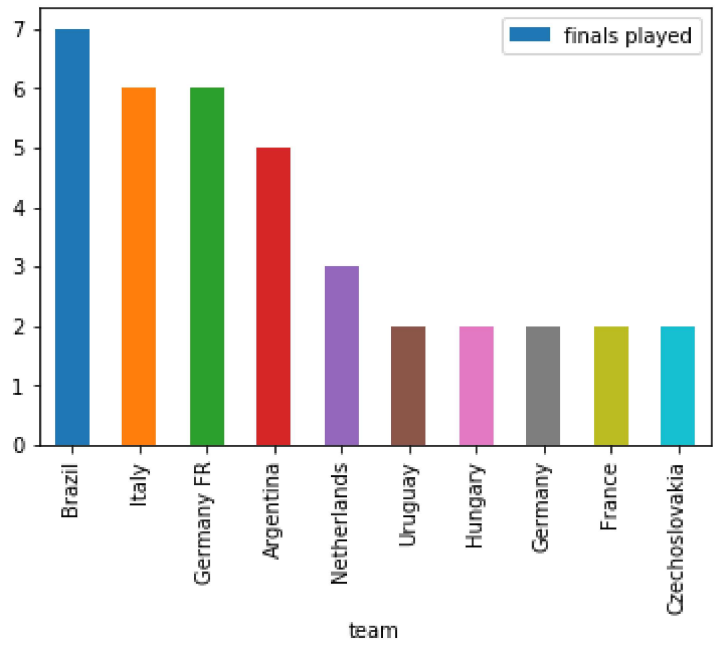


In [10]:

```
#top 10 team which played final most no. of times
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var3=pd.read_csv('WorldCups.csv')
var3=var3.fillna(0)
f=var3['Runners-Up'].value_counts()
f1=var3['Winner'].value_counts()
frames=[f,f1]
f2=pd.concat(frames)
f2=pd.DataFrame({'name':f2.index, 'final_appearence':f2.values})
f3=f2.groupby('name')['final_appearence'].sum()
f3=f3.sort_values(ascending=False)
f3=f3.head(10)
f3=pd.DataFrame({'team':f3.index, 'finals played':f3.values})
f3.plot.bar(x='team',y='finals played')
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218be5e92b0>
```

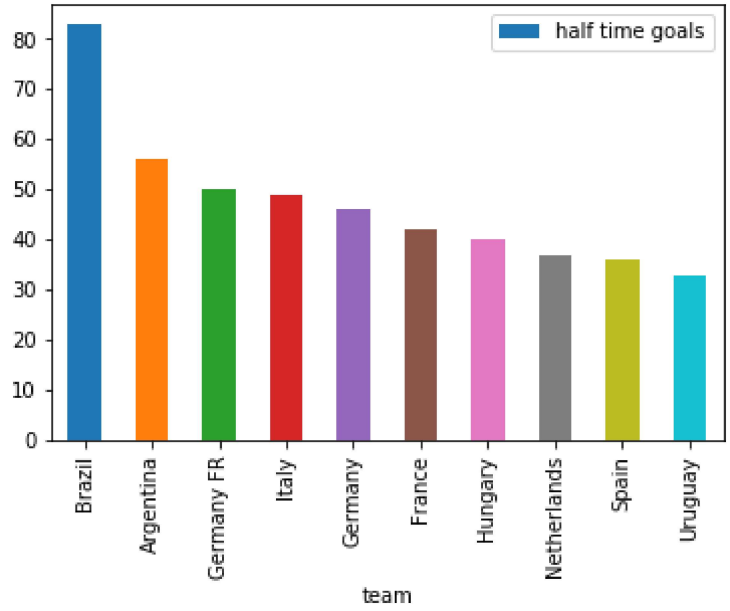


In [35]:

```
#top 10 team with most no. of halftime goals
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var0=pd.read_csv('WorldCupMatches.csv')
var0=var0.fillna(0)
d=var0.groupby('Home Team Name')['Half-time Home Goals'].sum()
d1=var0.groupby('Away Team Name')['Half-time Away Goals'].sum()
frames=[d,d1]
f5=pd.concat(frames)
f5=pd.DataFrame({'name':f5.index, 'half_time_goals':f5.values})
f6=f5.groupby('name')['half_time_goals'].sum()
f6=f6.sort_values(ascending=False)
f6=f6.head(10)
f6=pd.DataFrame({'team':f6.index, 'half time goals':f6.values})
f6.plot.bar(x='team',y='half time goals')
```

Out[35]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218bffb7518>
```

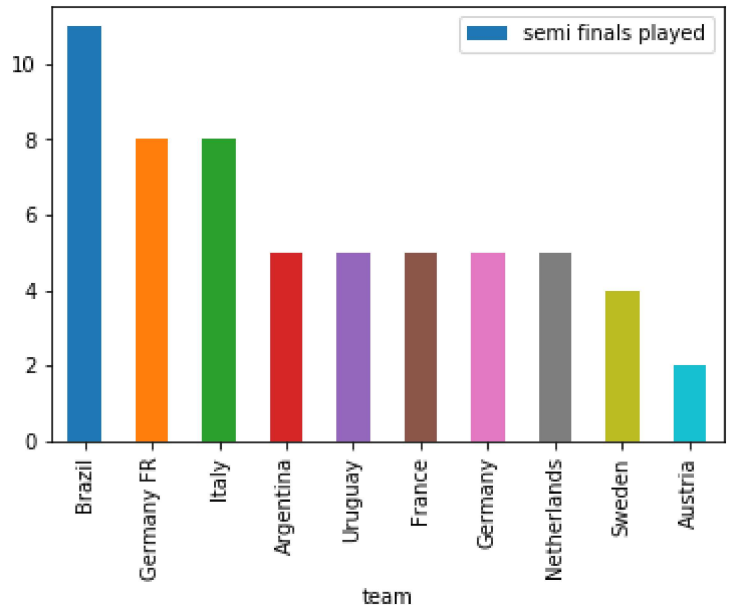


In [15]:

```
#top 10 teams which played semifinals most no. of times
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
var3=pd.read_csv('WorldCups.csv')
var3=var3.fillna(0)
f=var3['Runners-Up'].value_counts()
f1=var3['Winner'].value_counts()
fx=var3['Third'].value_counts()
fy=var3['Fourth'].value_counts()
frames=[f,f1,fx,fy]
f2=pd.concat(frames)
f2=pd.DataFrame({'name':f2.index, 'final_appearence':f2.values})
f3=f2.groupby('name')['final_appearence'].sum()
f3=f3.sort_values(ascending=False)
f3=f3.head(10)
f3=pd.DataFrame({'team':f3.index, 'semi finals played':f3.values})
f3.plot.bar(x='team',y='semi finals played')
```

Out[15]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218be9d6a58>
```

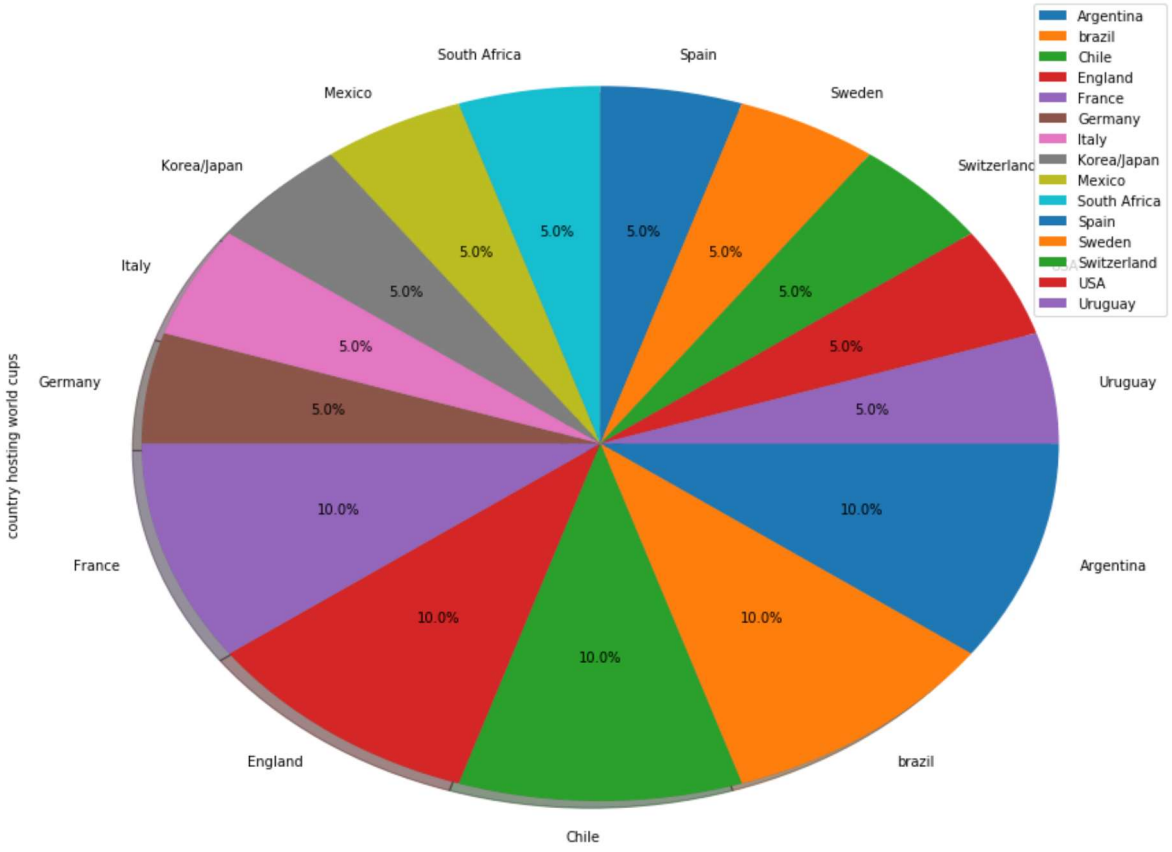


In [36]:

```
#country hosting highest no. of world cups
import numpy as np
import pandas as pd
import matplotlib as mpl
import matplotlib.pyplot as plt
abc=pd.read_csv('WorldCups.csv')
f=abc['Country'].value_counts()
l=['Argentina','brazil','Chile','England','France','Germany','Italy','Korea/Japan','Mexico'
f=pd.DataFrame({'countrry names':f.index, 'country hosting world cups':f.values})
plot1 = f.plot.pie(y='country hosting world cups',labels=l, autopct='%1.1f%%',counterclock=
plot1
```

Out[36]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x16abe8a7fd0>
```

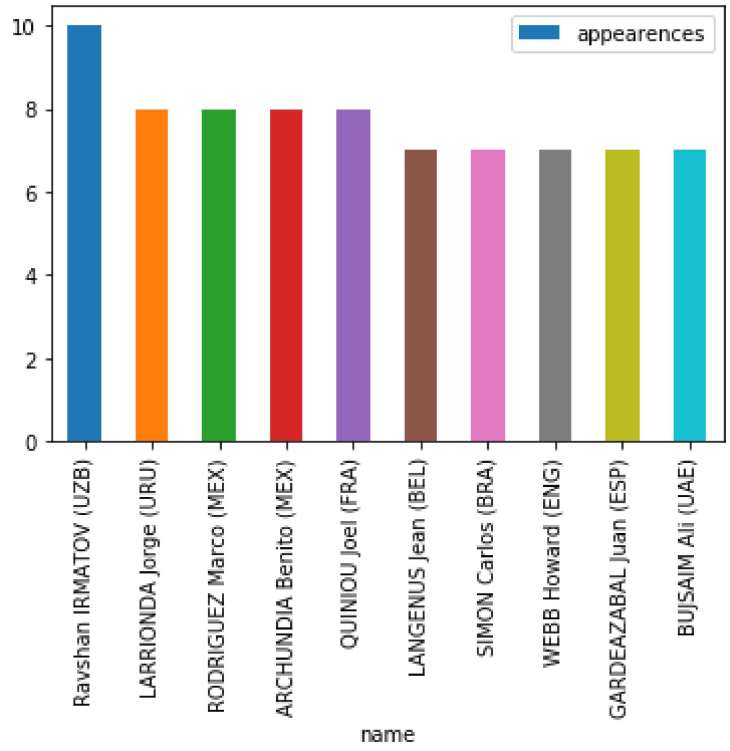


In [37]:

```
#top 10 referees with most no. of appearences
import numpy as np
import pandas as pd
import matplotlib as mlp
import matplotlib.pyplot as plt
k=pd.read_csv('WorldCupMatches.csv')
k1=k['Referee'].value_counts()
k1=k1.sort_values(ascending=False)
k1=k1.head(10)
k1=pd.DataFrame({'name':k1.index,'appearences':k1.values})
k1.plot.bar(x='name',y='appearences')
```

Out[37]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x16abe840160>
```

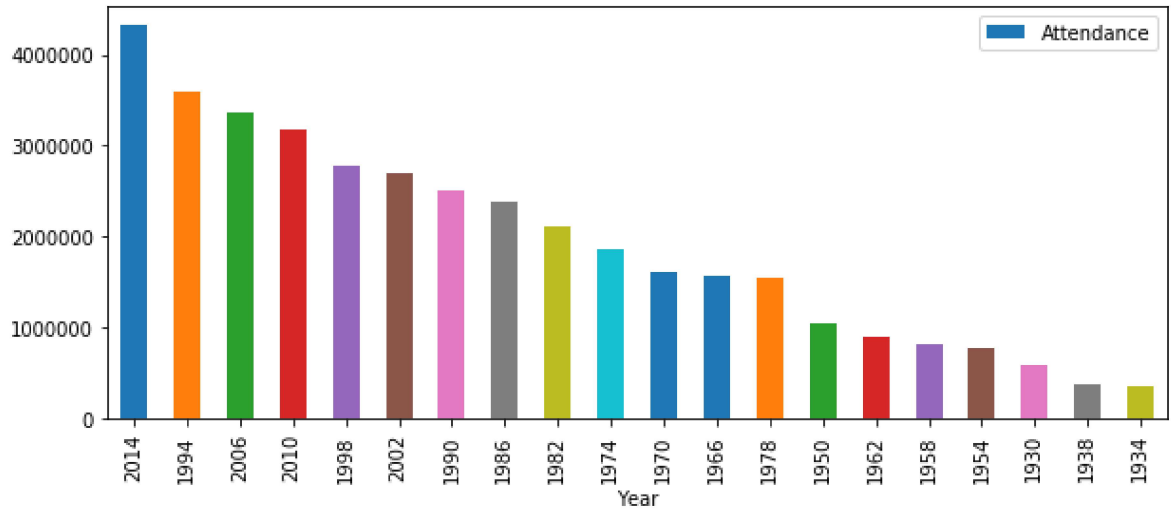


In [43]:

```
#World cup attendances in for each world cup in descending order
import numpy as np
import pandas as pd
import matplotlib as mlp
import matplotlib.pyplot as plt
k=pd.read_csv('WorldCupMatches.csv')
k=k.groupby('Year')['Attendance'].sum()
k=k.sort_values(ascending=False)
k=pd.DataFrame({'Year':k.index,'Attendance':k.values})
k.plot.bar(x='Year',y='Attendance',figsize=(10,4))
```

Out[43]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218c120b9e8>
```

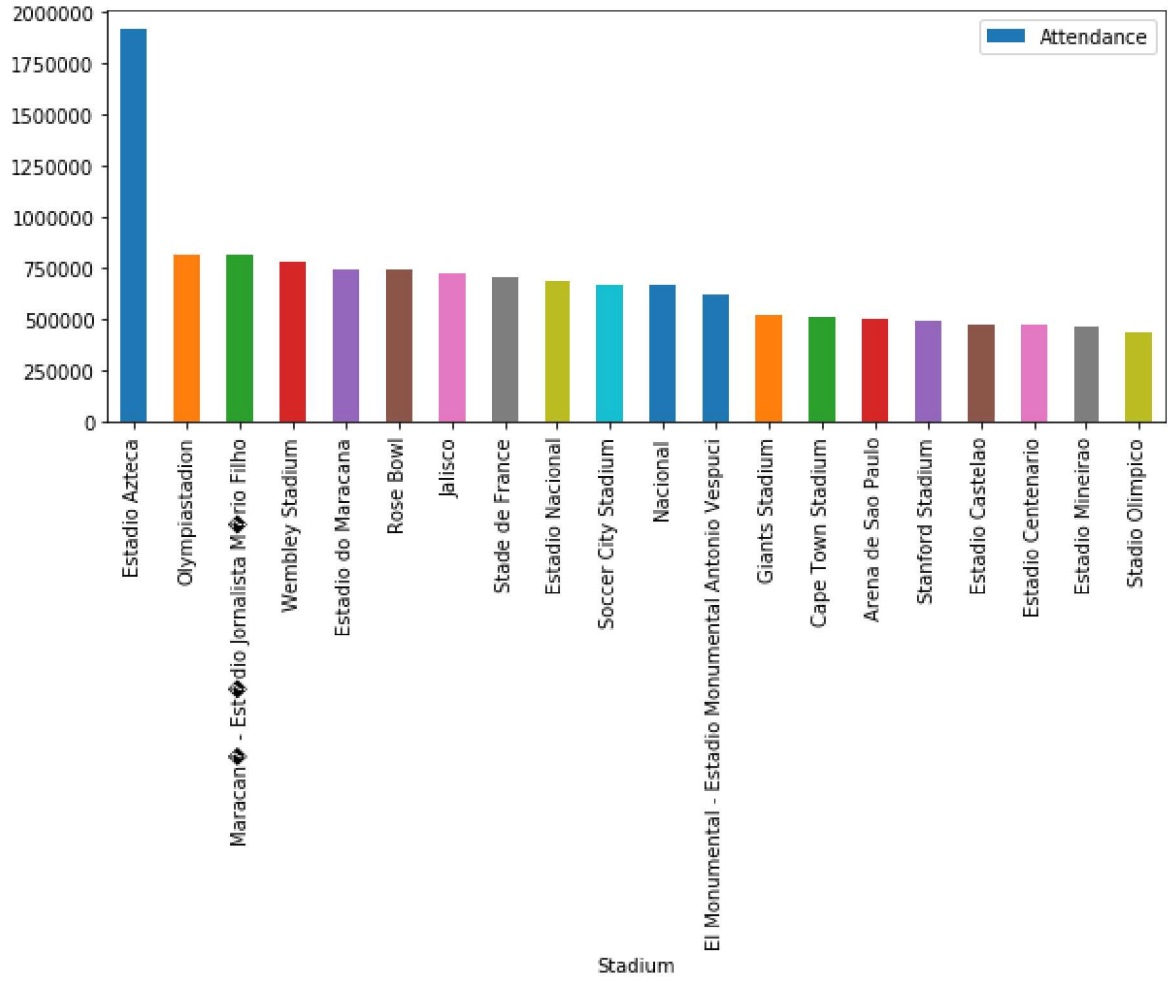


In [45]:

```
#top 20 stadiums based on attendance
import numpy as np
import pandas as pd
import matplotlib as mlp
import matplotlib.pyplot as plt
k=pd.read_csv('WorldCupMatches.csv')
k=k.groupby('Stadium')['Attendance'].sum()
k=k.sort_values(ascending=False)
k=k.head(20)
k=pd.DataFrame({'Stadium':k.index,'Attendance':k.values})
k.plot.bar(x='Stadium',y='Attendance',figsize=(10,4))
```

Out[45]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218c16749e8>
```

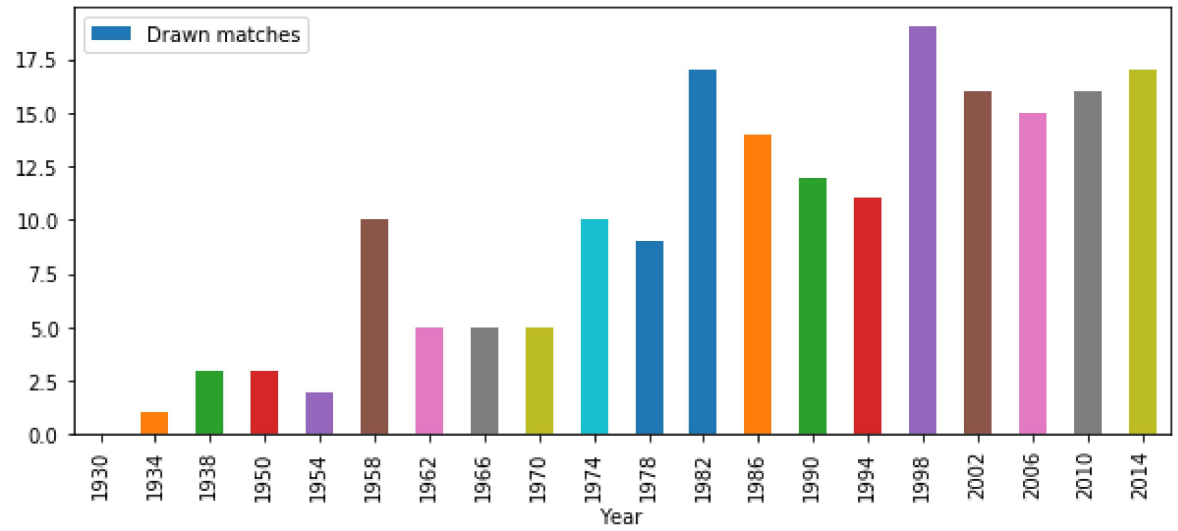


In [85]:

```
#matches drawn each year
import numpy as np
import pandas as pd
import matplotlib as mlp
import matplotlib.pyplot as plt
k=pd.read_csv('WorldCupMatches.csv')
k.fillna('abx')
k['new']=np.where(((k['Home Team Goals']==k['Away Team Goals'])),1,0)
k=k.groupby('Year')['new'].sum()
k=pd.DataFrame({'Year':k.index,'Drawn matches':k.values})
k.plot.bar(x='Year',y='Drawn matches',figsize=(10,4))
```

Out[85]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218c205d6a0>
```

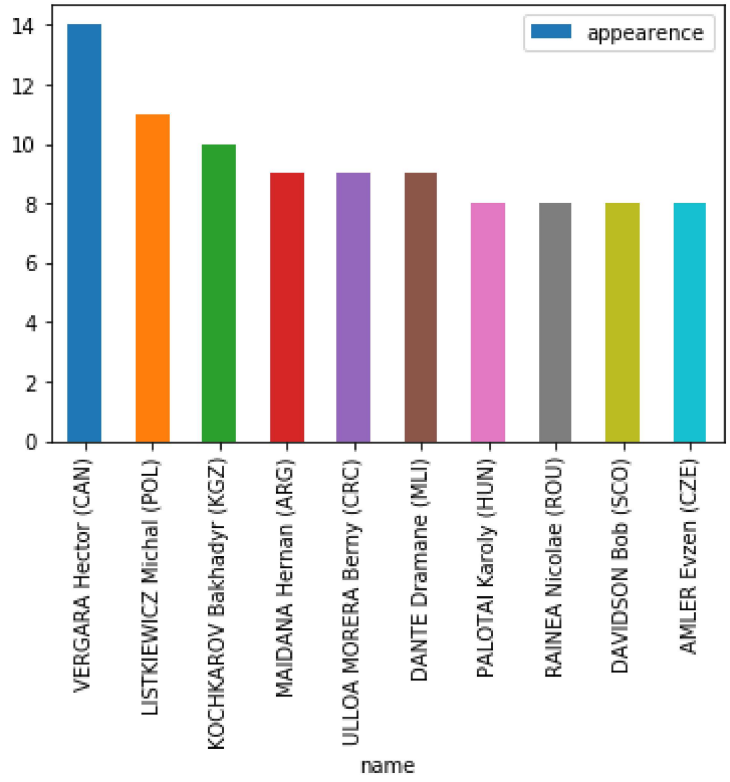


In [94]:

```
#most no. of appearences as assistant referee
import numpy as np
import pandas as pd
import matplotlib as mlp
import matplotlib.pyplot as plt
k=pd.read_csv('WorldCupMatches.csv')
k1=k['Assistant 1'].value_counts()
k2=k['Assistant 2'].value_counts()
fs=[k1,k2]
k3=pd.concat(fs)
k3=pd.DataFrame({'Name':k3.index,'appearence':k3.values})
k3=k3.groupby('Name')['appearence'].sum()
k3=k3.sort_values(ascending=False)
k3=k3.head(10)
k3=pd.DataFrame({'name':k3.index,'appearence':k3.values})
k3.plot.bar(x='name',y='appearence')
```

Out[94]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x218bfa64da0>
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

```
```