1. Loading main packages

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
%matplotlib inline
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
```

2. Path to all tables

In [2]:
```
Dept_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Department_Table.xlsx'
Educ_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Education_Table.xlsx'
Emp_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Employees_Table.xlsx'
Gend_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Gender_Table.xlsx'
JobR_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_JobRole_Table.xlsx'
MarStatus_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_MaritalStatus_Table.xlsx'
Sal_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Salaries_Table.xlsx'
Surv_T = r'C:\Users\Admin\Documents\Python Scripts\DABIS HR\DABIS_HR_DataSet\Dabis_Survey_Table.xlsx'
```

3. Exploring the Main table (Employees)

In [3]:
```
df = pd.read_excel(Emp_T)
df.head()
```

Out[3]:

	Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalC
0	1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
1	2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
2	3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
3	4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
4	5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

In [4]:
```
# Set EmployeeNumber as index
df = df.set_index('Employee_Id')
```

In [5]:
```
df.head()
```

Out[5]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

4. Joining all tables

Join 1 - Bringing Gender_table

In [6]:
```
df1 = pd.read_excel(Gend_T)
df1 = df1.set_index('Gender_Id')
df1.head()
```

Out[6]:

Gender_Id	Gender
1	Male
2	Female
3	Other

In [7]:
```
df2 = df.join(df1, on=['Gender_Id'],how='left',rsuffix='_GendT')
df2.head()
```

Out[7]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

Join 2 - Bringing Dept_table

In [8]:
```
df3 = pd.read_excel(Dept_T)
df3 = df3.set_index('Department_Id')
df3.head()
```

Out[8]:

Department_Id	Department
1	Research & Development
2	Sales
3	Human Resources
4	IT

In [9]:
```
# Employees + Gender + Department
df4 = df2.join(df3, on=['Department_Id'],how='left',rsuffix='_DepT')
df4.head()
```

Out[9]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

Join 3 - Bringing Educ_table

In [10]:
```
df5 = pd.read_excel(Educ_T)
df5 = df5.set_index('Education_Id')
df5.head()
```

Out[10]:

Education_Id	Education
1	High School
2	College
3	Bachelor's degree
4	Master's degree
5	PhD

In [11]:
```
# Employees + Gender + Department + Education
df6 = df4.join(df5, on=['Education_Id'],how='left',rsuffix='_EducT')
df6.head()
```

Out[11]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

Join 4 - Bringing JobRole_table

In [12]:
```
df7 = pd.read_excel(JobR_T)
df7 = df7.set_index('JobRole_Id')
df7.head()
```

Out[12]:

JobRole_Id	JobRole
1	Healthcare Representative
2	Laboratory Technician
3	Research Scientist
4	Manager
5	Manufacturing Director

In [13]:
```
# Employees + Gender + Department + Education + JobRole
df8 = df6.join(df7, on=['JobRole_Id'],how='left',rsuffix='_JobR_T')
df8.head()
```

Out[13]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

Join 5 - Bringing MarStatus_table

In [14]:
```
df9 = pd.read_excel(MarStatus_T)
df9 = df9.set_index('MartialStatus_Id')
df9.head()
```

Out[14]:

MartialStatus_Id	MartialStatus
1	Single
2	Married
3	Divorced

In [15]:
```
# Employees + Gender + Department + Education + JobRole + MartialStatus
df10 = df8.join(df9, on=['MaritalStatus_Id'],how='left',rsuffix='_MarStatusT')
df10.head()
```

Out[15]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

Join 6 - Bringing Salaries_table

In [16]:
```
df11 = pd.read_excel(Sal_T)
df11 = df11.set_index('Employee_Id')
df11.head()
```

Out[16]:

Employee_Id	HourlyRate
1	94
2	61
3	92
4	56
5	40

In [17]:
```
df12 = df10.join(df11, on=['Employee_Id'],how='left',rsuffix='_SalT')
df12.head()
```

Out[17]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

5 rows × 21 columns

Join 7 - Bringing Survey_table

In [18]:
```
df13 = pd.read_excel(Surv_T)
df13 = df13.set_index('Employee_Id')
df13.head()
```

Out[18]:

Employee_Id	NumCompaniesWorked	PerformanceRating	TrainingTimesLastYear	YearsAtCompany
1	8	3	0	6
2	1	4	3	10
3	6	3	3	0
4	1	3	3	8
5	9	3	3	2

In [19]:
```
fullTable = df12.join(df13, on=['Employee_Id'],how='left',rsuffix='_SurvT')
fullTable.head()
```

Out[19]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

5 rows × 31 columns

5. Finale Table with all relationships

In [22]:
```
fullTable.head()
```

Out[22]:

Employee_Id	Title	FirstName	LastName	BirthDate	EmailAddress	TotalChil
1	Mr.	Melody	Goel	1966-04-08	melody.goel@dabis.consulting.ca	
2	Mr.	Paige	Diaz	1965-05-14	paige.diaz@dabis.consulting.ca	
3	Mr.	Annabelle	Martin	1965-08-12	annabelle.martin@dabis.consulting.ca	
4	Ms.	David	Ye	1968-02-15	david.ye@dabis.consulting.ca	
5	Ms.	Leroy	Raman	1968-08-08	leroy.raman@dabis.consulting.ca	

5 rows × 31 columns

In [20]:
```
print(f"fullTable data shape {fullTable.shape}")
```

```
fullTable data shape (1470, 31)
```

In [21]:
```
fullTable.isnull().sum()
```

Out[21]:
```
Title                     0
FirstName                 0
LastName                  0
BirthDate                 0
EmailAddress              0
TotalChildren             0
Department_Id             0
Education_Id              0
Gender_Id                 0
Age                       0
JobLevel                  0
JobRole_Id                0
MaritalStatus_Id          0
YearofHire                0
DateofHire                0
Gender                    0
Department                0
Education                 0
JobRole                   0
MartialStatus             0
HourlyRate                0
NumCompaniesWorked        0
PerformanceRating         0
TrainingTimesLastYear     0
YearsAtCompany            0
YearsInCurrentRole        0
YearsSinceLastPromotion   0
YearsWithCurr4            0
JobSatisfaction           0
EnvironmentSatisfaction   0
WorkLifeBalance           0
dtype: int64
```

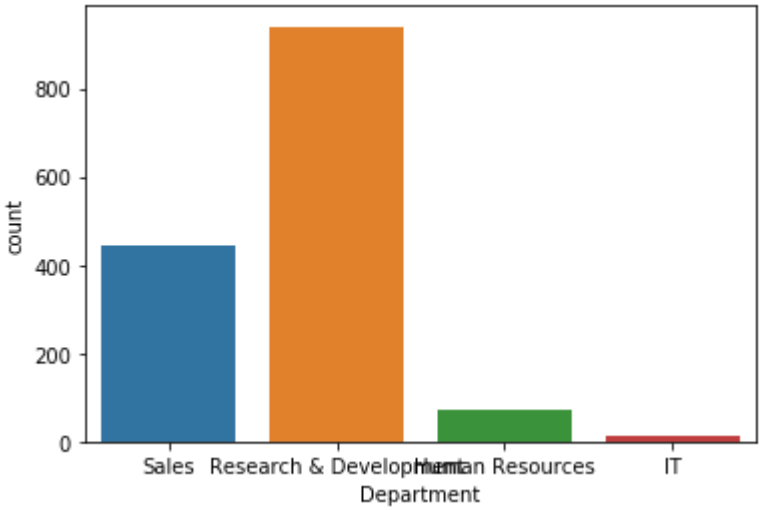
6. Visualization

A. Number of Employees by Department

In [25]:
```
fullTable.groupby('Department')['Department'].agg("count")
```

Out[25]:
```
Department
Human Resources            73
IT                         14
Research & Development    940
Sales                     443
Name: Department, dtype: int64
```

In [26]:
```
sns.countplot(x=fullTable["Department"], data=fullTable)
plt.savefig("Image_NumberOfEmployeesByDepartments.png")
```

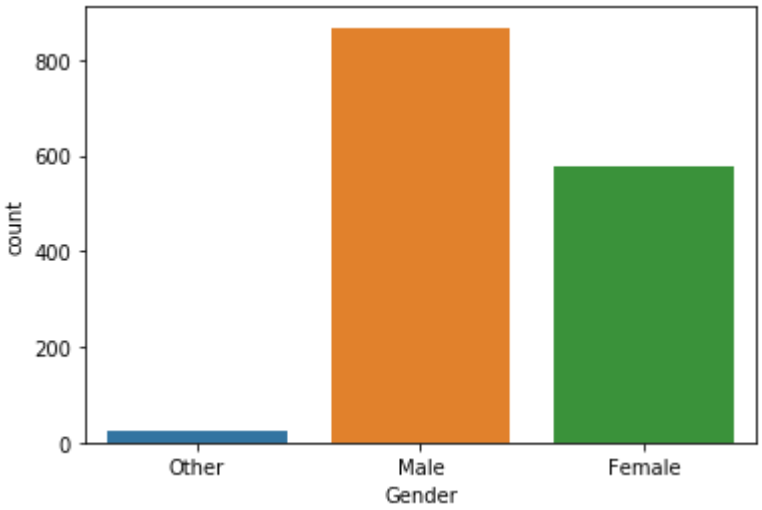


B. Number of Employees by Gender

In [30]:
```
fullTable.groupby('Gender')['Gender'].agg("count")
```

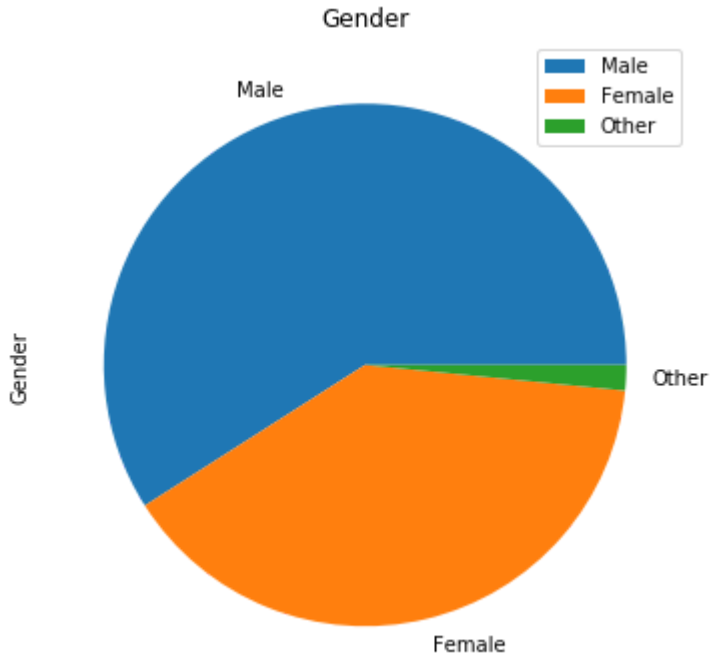
Out[30]:
```
Gender
Female    579
Male      868
Other      23
Name: Gender, dtype: int64
```

In [31]:
```
sns.countplot(x=fullTable["Gender"], data=fullTable)
plt.savefig("Image_NumberOfEmployeesByGender.png")
```



In [33]:
```
GenderVisual = fullTable['Gender'].value_counts()
```

In [34]:
```
GenderVisual.plot(kind='pie', title='Gender', figsize=(6,6))
plt.legend()
plt.show()
```



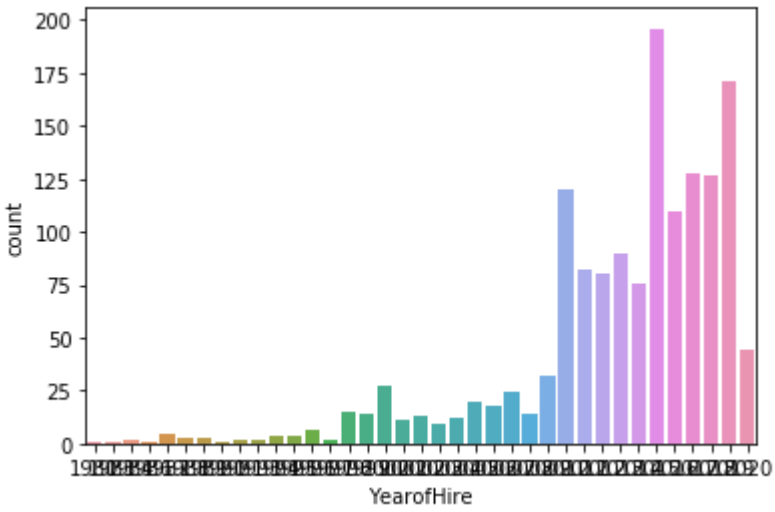
C. Number of Employees by Gender in each Department

In [32]:
```
fullTable.groupby(['Department','Gender'])['Gender'].agg("count")
```

Out[32]:
```
Department              Gender
Human Resources         Female     27
                        Male       45
                        Other       1
IT                      Female      6
                        Male        7
                        Other       1
Research & Development  Female    361
                        Male      565
                        Other      14
Sales                   Female    185
                        Male      251
                        Other       7
Name: Gender, dtype: int64
```

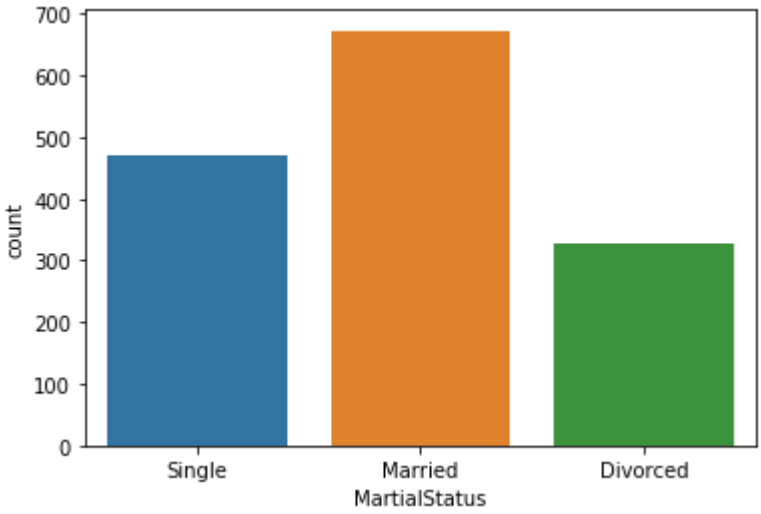
C. Year of Hire

In [36]:
```
sns.countplot(x=fullTable['YearofHire'], data=fullTable)
plt.savefig("Image_YearofHire.png")
```



D. EMployees by Marital Status

In [44]:
```
sns.countplot(x=fullTable["MartialStatus"], data=fullTable)
plt.savefig("Image_EmployeesByMaritalStatus.png")
```



E. Average Calculation

In [39]:
```
AtCompany = fullTable.YearsAtCompany.mean()
print(f"Average of years where our employees work at the company is {AtCompan
y}")
```

```
Average of years where our employees work at the company is 7.008163265306122
5
```

In [40]:
```
InCurrentRole = fullTable.YearsInCurrentRole.mean()
print(f"Average of years where our employees work in current role is {InCurren
tRole}")
```

```
Average of years where our employees work in current role is 4.22925170068027
2
```

In [41]:
```
CompaniesWorked = fullTable.NumCompaniesWorked.mean()
print(f"Average of number of companies where our employees worked in the past
 is {CompaniesWorked}")
```

```
Average of number of companies where our employees worked in the past is 2.69
31972789115646
```