





## 1 Preparing Data

## 2 Visualizing Data

- Box Plot
- Bar Plot
- Histogram
- Kernel Density Estimate (KDE) Plot





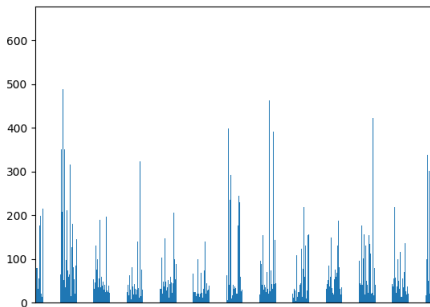






# Visualizing the bar plot

```
import matplotlib.pyplot as plt
ED_ManHour.plot.bar()
plt.xticks([])
plt.show()
```

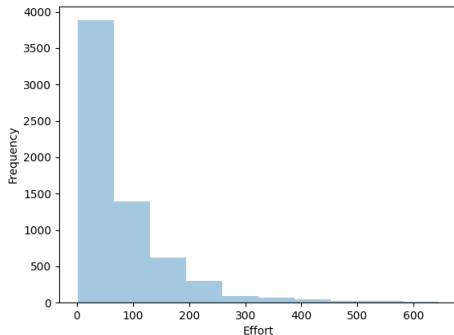


# Visualizing the histogram

The histogram highlights the discrete probability density distribution (frequency distribution).

# Visualizing the histogram

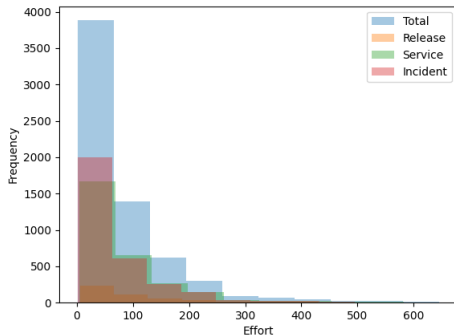
```
import matplotlib.pyplot as plt
ED_ManHour.plot.hist(bins=10, alpha=0.4)
plt.xlabel('ManHour')
plt.show()
```



## Visualizing the histogram – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.hist(bins=10, alpha=0.4)
ED1 = ED.loc[ED["ITILProcessType"]=="Release Management"]
ED1_ManHour = ED1.loc[:, "ManHour"]
ED1_ManHour.plot.hist(bins=10, alpha=0.4)
ED2 = ED.loc[ED["ITILProcessType"]=="Service Request"]
ED2_ManHour = ED2.loc[:, "ManHour"]
ED2_ManHour.plot.hist(bins=10, alpha=0.4)
ED3 = ED.loc[ED["ITILProcessType"]=="Incident Management"]
ED3_ManHour = ED3.loc[:, "ManHour"]
ED3_ManHour.plot.hist(bins=10, alpha=0.4)
plt.legend(['Total', 'Release', 'Service', 'Incident'])
plt.xlabel('ManHour')
plt.show()
```

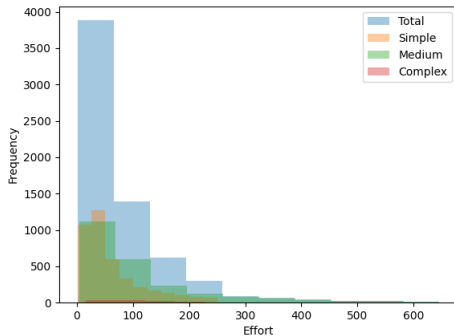
# Visualizing the histogram – Let's go deeper in the data



## Visualizing the histogram – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.hist(bins=10, alpha=0.4)
ED4 = ED.loc[ED["Complexity"]=="Simple"]
ED4_ManHour = ED4.loc[:, "ManHour"]
ED4_ManHour.plot.hist(bins=10, alpha=0.4)
ED5 = ED.loc[ED["Complexity"]=="Medium"]
ED5_ManHour = ED5.loc[:, "ManHour"]
ED5_ManHour.plot.hist(bins=10, alpha=0.4)
ED6 = ED.loc[ED["Complexity"]=="Complex"]
ED6_ManHour = ED6.loc[:, "ManHour"]
ED6_ManHour.plot.hist(bins=10, alpha=0.4)
plt.legend(['Total', 'Simple', 'Medium', 'Complex'])
plt.xlabel('ManHour')
plt.show()
```

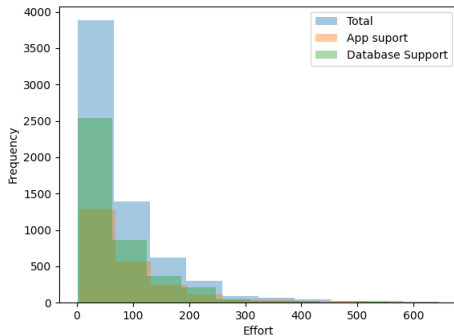
# Visualizing the histogram – Let's go deeper in the data



## Visualizing the histogram – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.hist(bins=10, alpha=0.4)
ED7 = ED.loc[ED["Service Domain type"]=="App Support"]
ED7_ManHour = ED7.loc[:, "ManHour"]
ED7_ManHour.plot.hist(bins=10, alpha=0.4)
ED8 = ED.loc[ED["Service Domain type"]=="Database Support"]
ED8_ManHour = ED8.loc[:, "ManHour"]
ED8_ManHour.plot.hist(bins=10, alpha=0.4)
plt.legend(['Total', 'App Support', 'Database Support'])
plt.xlabel('ManHour')
plt.show()
```

# Visualizing the histogram – Let's go deeper in the data

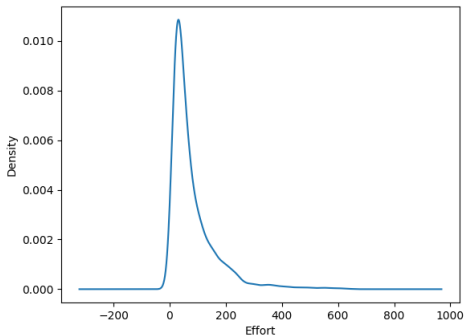


# Visualizing the kernel density estimate (KDE) plot

The kernel density estimate (KDE) plot highlights the continuous probability density distribution. This is analogous to a histogram in continuous space.

# Visualizing the kernel density estimate (KDE) plot

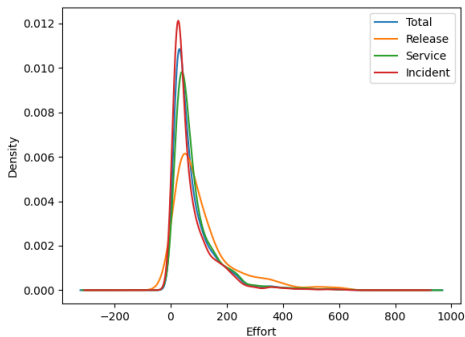
```
import matplotlib.pyplot as plt
ED_ManHour.plot.kde()
plt.xlabel('ManHour')
plt.show()
```



## Visualizing the KDE plot – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.kde()
ED1 = ED.loc[ED["ITILProcessType"]=="Release Management"]
ED1_ManHour = ED1.loc[:, "ManHour"]
ED1_ManHour.plot.kde()
ED2 = ED.loc[ED["ITILProcessType"]=="Service Request"]
ED2_ManHour = ED2.loc[:, "ManHour"]
ED2_ManHour.plot.kde()
ED3 = ED.loc[ED["ITILProcessType"]=="Incident Management"]
ED3_ManHour = ED3.loc[:, "ManHour"]
ED3_ManHour.plot.kde()
plt.legend(['Total', 'Release', 'Service', 'Incident'])
plt.xlabel('ManHour')
plt.show()
```

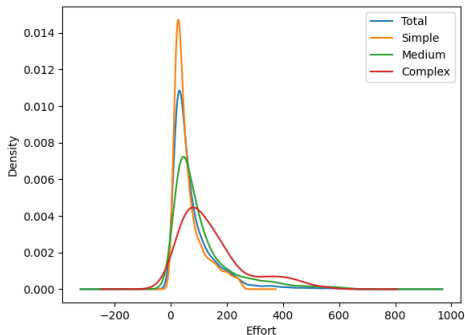
# Visualizing the KDE plot – Let's go deeper in the data



## Visualizing the KDE plot – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.kde()
ED4 = ED.loc[ED["Complexity"]=="Simple"]
ED4_ManHour = ED4.loc[:, "ManHour"]
ED4_ManHour.plot.kde()
ED5 = ED.loc[ED["Complexity"]=="Medium"]
ED5_ManHour = ED5.loc[:, "ManHour"]
ED5_ManHour.plot.kde()
ED6 = ED.loc[ED["Complexity"]=="Complex"]
ED6_ManHour = ED6.loc[:, "ManHour"]
ED6_ManHour.plot.kde()
plt.legend(['Total', 'Simple', 'Medium', 'Complex'])
plt.xlabel('ManHour')
plt.show()
```

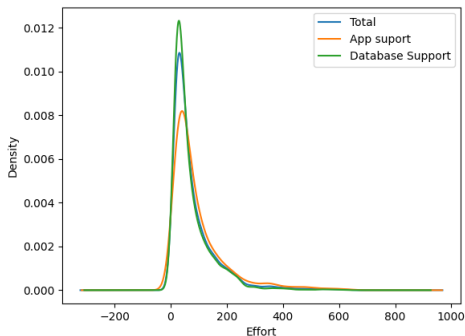
# Visualizing the KDE plot – Let's go deeper in the data



# Visualizing the KDE plot – Let's go deeper in the data

```
import matplotlib.pyplot as plt
ED_ManHour.plot.kde()
ED7 = ED.loc[ED["Service Domain type"]=="App Support"]
ED7_ManHour = ED7.loc[:, "ManHour"]
ED7_ManHour.plot.kde()
ED8 = ED.loc[ED["Service Domain type"]=="Database Support"]
ED8_ManHour = ED8.loc[:, "ManHour"]
ED8_ManHour.plot.kde()
plt.legend(['Total', 'App Support', 'Database Support'])
plt.xlabel('ManHour')
plt.show()
```

# Visualizing the KDE plot – Let's go deeper in the data



# Problems

- 1 Create the KDE plots for all the 18 combinations of tasks and overlay them on a single figure.