

October 8, 2021

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[2]: df = pd.read_csv('microsoft-malware-prediction/train.csv')
df1 = df[['Census_InternalPrimaryDiagonalDisplaySizeInInches',
↪ 'Census_InternalPrimaryDisplayResolutionHorizontal',
↪ 'Census_InternalPrimaryDisplayResolutionVertical', 'Wdft_IsGamer',
↪ 'HasDetections']]
df1.dropna(how = 'any', inplace = True)
```

C:\Users\Gary\AppData\Roaming\Python\Python38\site-packages\IPython\core\interactiveshell.py:3146: DtypeWarning: Columns (28) have mixed types.Specify dtype option on import or set low\_memory=False.

has\_raised = await self.run\_ast\_nodes(code\_ast.body, cell\_name, <ipython-input-2-6270d84cdf44>:3: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df1.dropna(how = 'any', inplace = True)
```

```
[3]: fig, ax = plt.subplots(figsize = (8, 6))
mask = np.triu(np.ones_like(df1.corr(), dtype = np.bool))[1:, :-1]
xlabel = ['Diagonal Size', 'Horizontal Display', 'Vertical Display', 'Is_Gamer']
ylabel = ['Horizontal Display', 'Vertical Display', 'Is_Gamer', 'Malware_
↪ Detection']
sns.heatmap(df1.corr().iloc[1:, :-1], vmin = -1, vmax = 1, cmap = 'Blues', annot_
↪ = True, fmt = ".2f", linewidth = 0.3, xticklabels = xlabel, yticklabels =
↪ ylabel, mask = mask)
plt.title('Correlation Heatmap')
plt.yticks(rotation = 0)
```

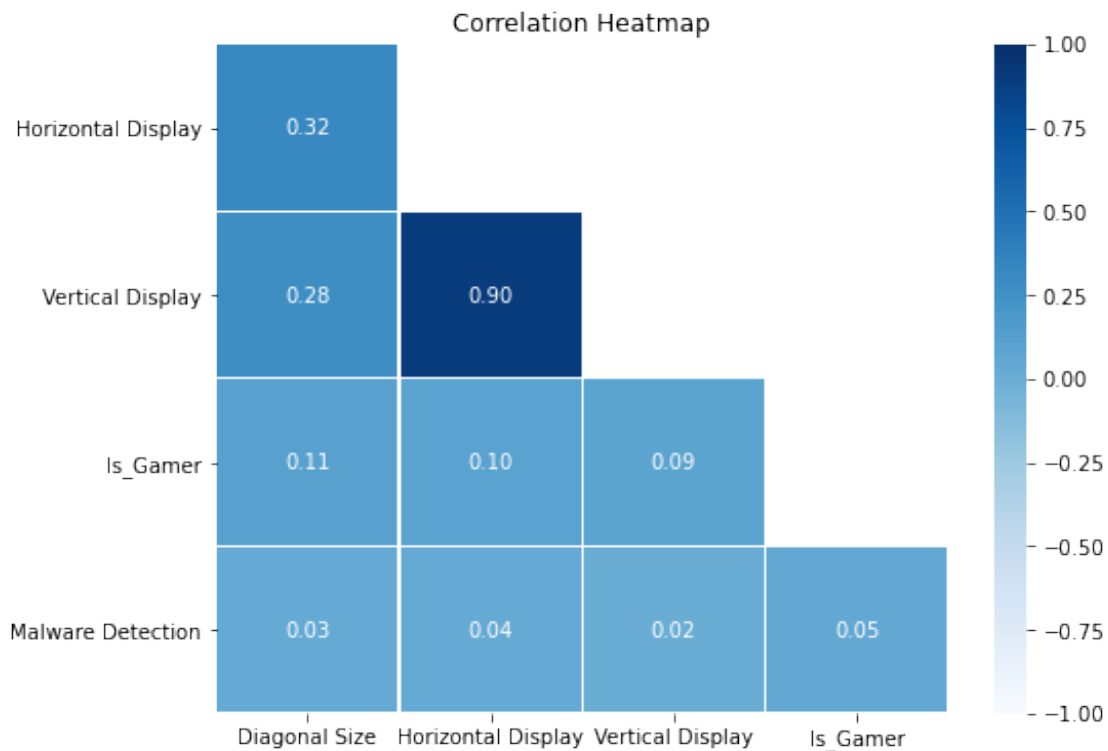
<ipython-input-3-afa71596a192>:2: DeprecationWarning: `np.bool` is a deprecated alias for the builtin `bool`. To silence this warning, use `bool` by itself. Doing this will not modify any behavior and is safe. If you specifically wanted the numpy scalar type, use `np.bool\_` here.

Deprecated in NumPy 1.20; for more details and guidance:

<https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations>

```
mask = np.triu(np.ones_like(df1.corr(), dtype = np.bool))[1:, :-1]
```

```
[3]: (array([0.5, 1.5, 2.5, 3.5]),  
      [Text(0, 0.5, 'Horizontal Display'),  
       Text(0, 1.5, 'Vertical Display'),  
       Text(0, 2.5, 'Is_Gamer'),  
       Text(0, 3.5, 'Malware Detection')])
```



```
[4]: fig, ax = plt.subplots(figsize = (8, 6))  
sns.boxplot(x = 'Wdft_IsGamer', y =  
            ↪ 'Census_InternalPrimaryDisplayResolutionHorizontal', data = df1).set(xlabel=  
            ↪ 'Is_Gamer', ylabel = 'Horizontal Display')  
ax.set_title('Boxplot of gamer/non-gamer versus horizontal display')
```

```
[4]: Text(0.5, 1.0, 'Boxplot of gamer/non-gamer versus horizontal display')
```

