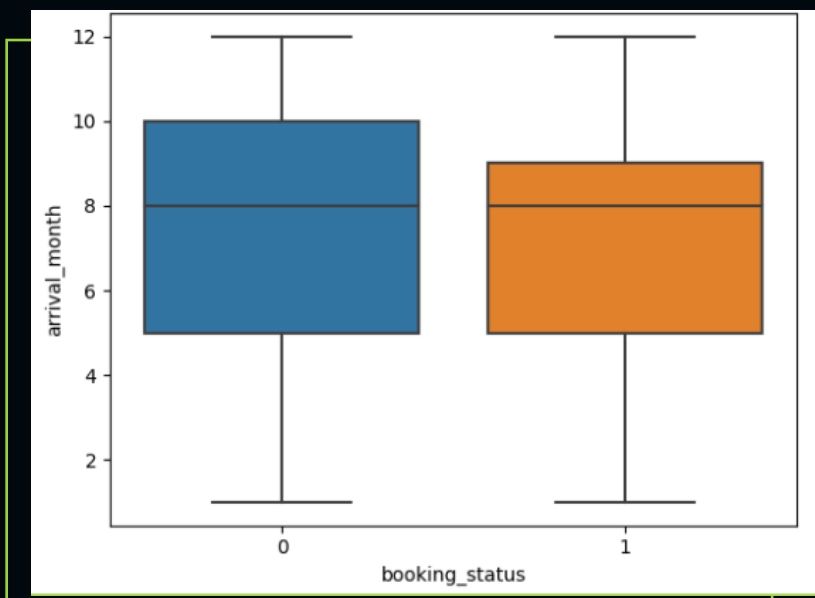


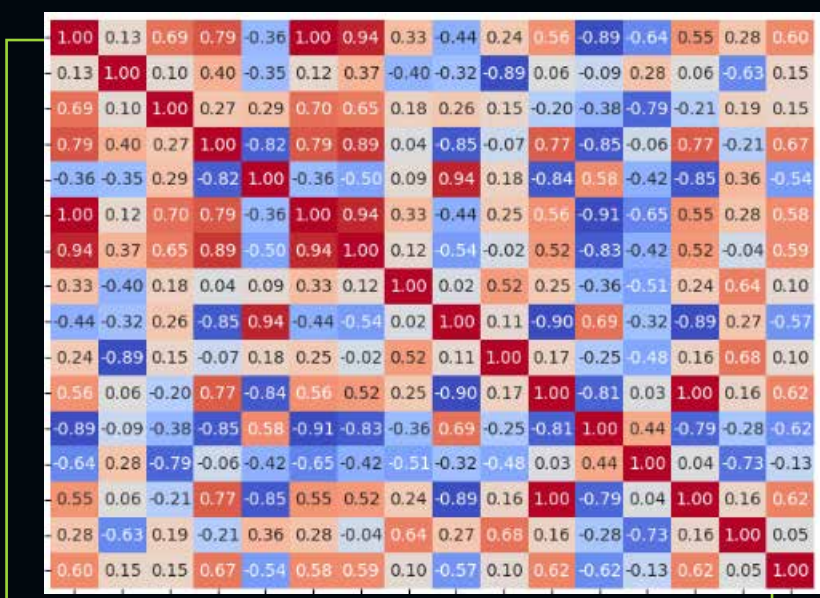
_ANÁLISE EXPLORATÓRIA

#A análise dos dados pode nos ajudar a determinar quais features nos permitem diferenciar melhor as diferentes classes do problema. Porém devemos ter cuidado com a inspeção visual já que é possível pequenas diferenças entre determinadas variáveis sejam importantes para alguns modelos de machine learning.

#Algumas ferramentas que podemos usar são os boxplots e a matriz de correlação.



#Boxplots



#Matriz de correlação

_MÉTODOS DE SELEÇÃO DE FEATURES

```
> from sklearn.feature_selection import SelectFromModel, RFE, RFECV
```

#SelectFromModel



#A função SelectFromModel é uma técnica usada para selecionar características (ou features) importantes a partir de um modelo de aprendizado de máquina já treinado. Essa seleção de características é útil para melhorar o desempenho do modelo, reduzir a dimensionalidade dos dados e eliminar características menos relevantes.

Seleção das features

```
> modelo_fselecionadas = SelectFromModel(modelo, max_features=10, prefit=True)
```

Filtragem das features selecionadas

```
> X_treino_selecionado = modelo_fselecionadas.transform(X_treino);  
> X_teste_selecionado = modelo_fselecionadas.transform(X_teste);
```

#RFE



#A Recursive Feature Elimination (RFE) funciona eliminando recursivamente as características menos importantes com base em um modelo pré-definido, permitindo que você identifique um subconjunto ideal de características que maximizam o desempenho do modelo. A RFE é particularmente útil quando você deseja reduzir a dimensionalidade dos dados e manter apenas as características mais relevantes, melhorando assim a eficiência e a capacidade preditiva do seu modelo.

#Instanciamento do selecionador

```
> selecionador = RFE(modelo,n_features_to_select=5, step=1)
```

#Ajuste aos dados de treino

```
> selecionador.fit(X_treino, y_treino)
```

#Filtragem das features selecionadas

```
> X_treino_selecionado = selecionador.transform(X_treino)  
> X_teste_selecionado = selecionador.transform(X_teste)
```

#RFECV



#A Recursive Feature Elimination with Cross-Validation (RFECV) é uma extensão da RFE no Scikit-learn que inclui validação cruzada. Essa técnica automatizada seleciona as melhores features para um modelo de aprendizado de máquina, mas também avalia sua performance usando validação cruzada, garantindo que a seleção seja robusta. Ela realiza a eliminação recursiva de características, avaliando cada conjunto de características por meio de validação cruzada, permitindo que você encontre não apenas as melhores características, mas também a quantidade ideal delas para otimizar o desempenho do seu modelo. A RFECV é especialmente útil para evitar overfitting e garantir que a seleção de características seja estável em diferentes divisões dos dados.

Instanciamento do StratifiedKfold

```
> cv = StratifiedKfold(5,shuffle=True,random_state=500)
```

Instanciamento do RFECV

```
> rfecv = RFECV(  
>     estimator=modelo,  
>     step=1,  
>     cv=cv,  
>     scoring='roc_auc',  
>     min_features_to_select=1,  
>     importance_getter='auto'  
> )  
> rfecv.fit(X_treino,y_treino)
```



*clique nos ícones para **acessar** as respectivas **documentações***