

---

# Sessão 05

WiDS Recife Live Coding, 18/01/2020

---

# Como serão os live codings?

- Sessões ao-vivo todos os sábados das 14h às 15h
  - Código e slides serão disponibilizados no nosso site
  - O objetivo é treinar para participar do [Datathon](#) em 2020
-

---

# Roteiro

- Análise exploratória dos dados do [Datathon 2020](#) com o Seaborn

# O problema

- O objetivo é criar um modelo que preveja a probabilidade do paciente sobreviver a partir dos dados das primeiras 24 horas dele na UTI do hospital
- A métrica utilizada para avaliação é a ROC AUC

# Etapas para resolver o problema

1. Importar os dados
2. **Explorar os dados**
3. Treinar o modelo com o conjunto de treinamento
4. Avaliar o modelo com o conjunto de testes

---

# Explorando os dados

---

# Explorando os dados

- Vamos usar o Seaborn
  - É uma biblioteca em Python que usa o Matplotlib para criar as visualizações
-

---

# Vendo os tipos dos dados

```
df.dtypes.value_counts()
```

```
float64    170
```

```
object      8
```

```
int64       8
```

---



---

# Como o Seaborn funciona

- Cada tipo de gráfico possui um método específico
- Podemos customizar o gráfico passando parâmetros
- Exemplo:
  - Gráfico de barra com o número de ocorrências de cada valor

```
import seaborn as sns
```

```
sns.countplot(x=*nome da coluna*, data=*nome do  
dataframe*)
```

---

---

# Customizando gráficos

- Podemos customizar os gráficos usando funções do Matplotlib
- Exemplo:
  - Rotacionar as labels do gráfico

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(x=*nome da coluna*, data=*nome do
dataframe*)
plt.xticks(rotate=90)
```

---

---

# Heatmap

```
sns.heatmap(pd.crosstab(df.ethnicity,  
df.hospital_death, margins=True,  
normalize="index"),annot=True)
```

---

---

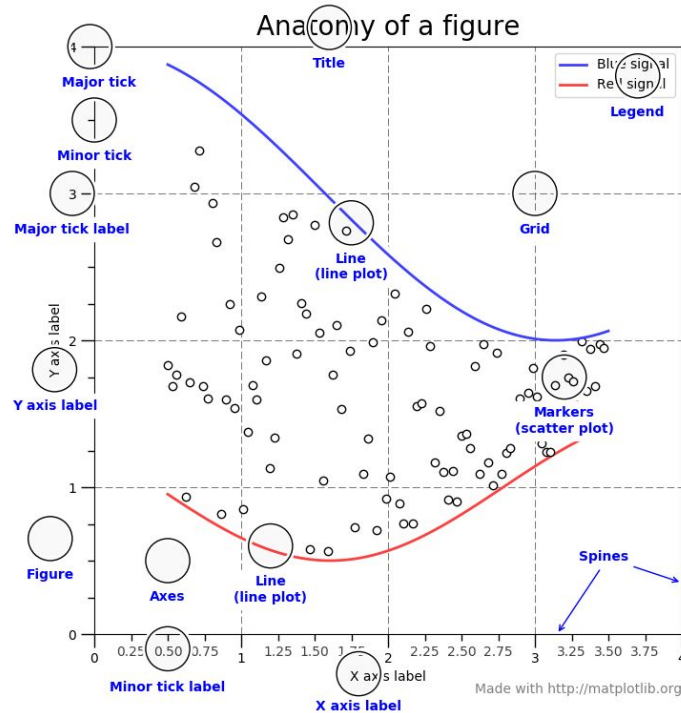
# Boxplot

```
sns.boxplot(data=df.select_dtypes("float").iloc[:,[0,1,2,3,4,5]])
```

```
plt.xticks(rotation=90)
```

---

# Anatomia de um gráfico



Fonte:

<https://matplotlib.org/3.1.1/tutorials/introductory/usage.html>  
[#sphx-glr-tutorials-introductory-usage-py](#)

---

# Competição no kaggle

---

# Competição no kaggle

- Normalmente as pessoas compartilham notebooks e discutem os resultados/melhorias
    - <https://www.kaggle.com/c/widsdatathon2020/notebooks>
  - Podemos usar estes notebooks para não reinventar a roda
    - <https://www.kaggle.com/jayjay75/wids2020-lgb-starter-script>
-

---

# Competição no kaggle

- Dicionário dos dados
  - <https://www.kaggle.com/c/widsdatathon2020/data>



---

# Obrigada!

E até semana que vem!

---

# —

## Referências

- [1] <https://www.kaggle.com/c/widsdatathon2020/overview/description>
- [2] <https://elitedatascience.com/python-seaborn-tutorial>
- [3] [https://en.wikipedia.org/wiki/APACHE II](https://en.wikipedia.org/wiki/APACHE_II)