



Fábio Mori

Programação Aplicada

IMERSÃO NO
MOTORSPORT
PARA MULHERES

A MATTHEIS
MOTORSPORT

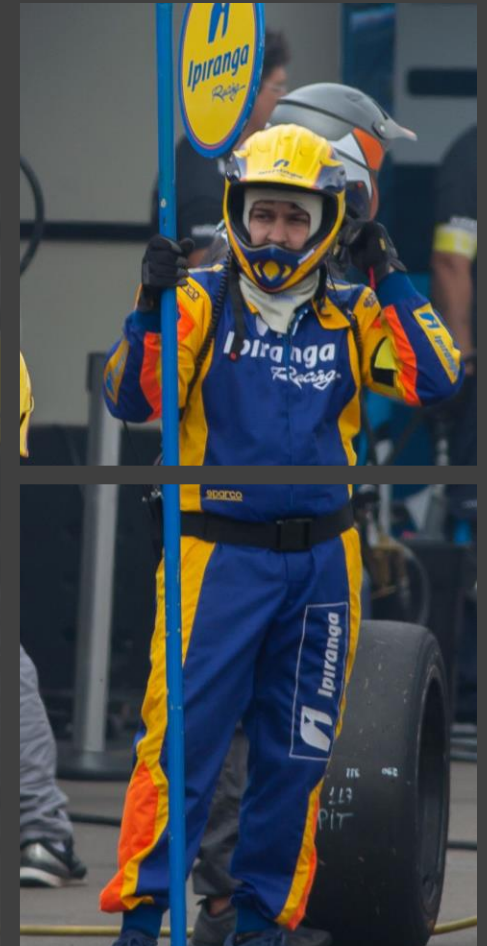
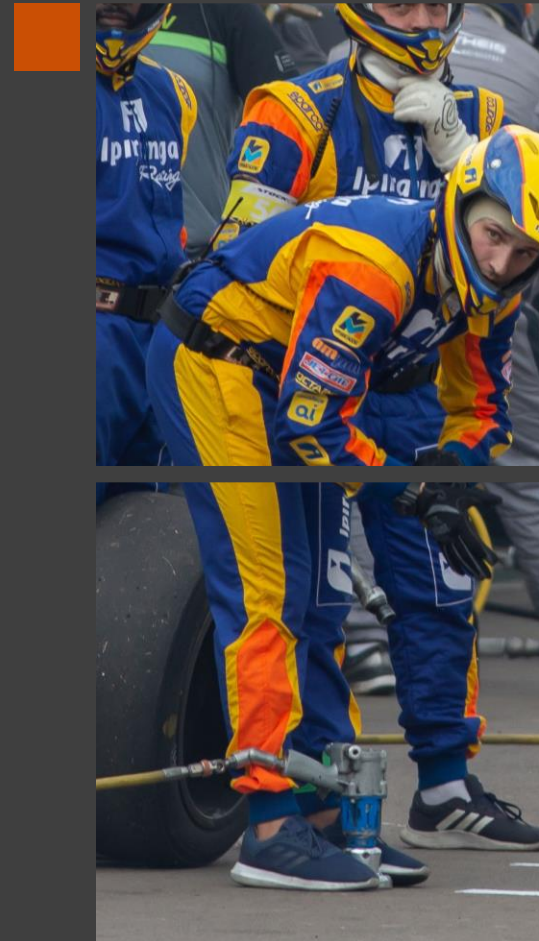
Agenda

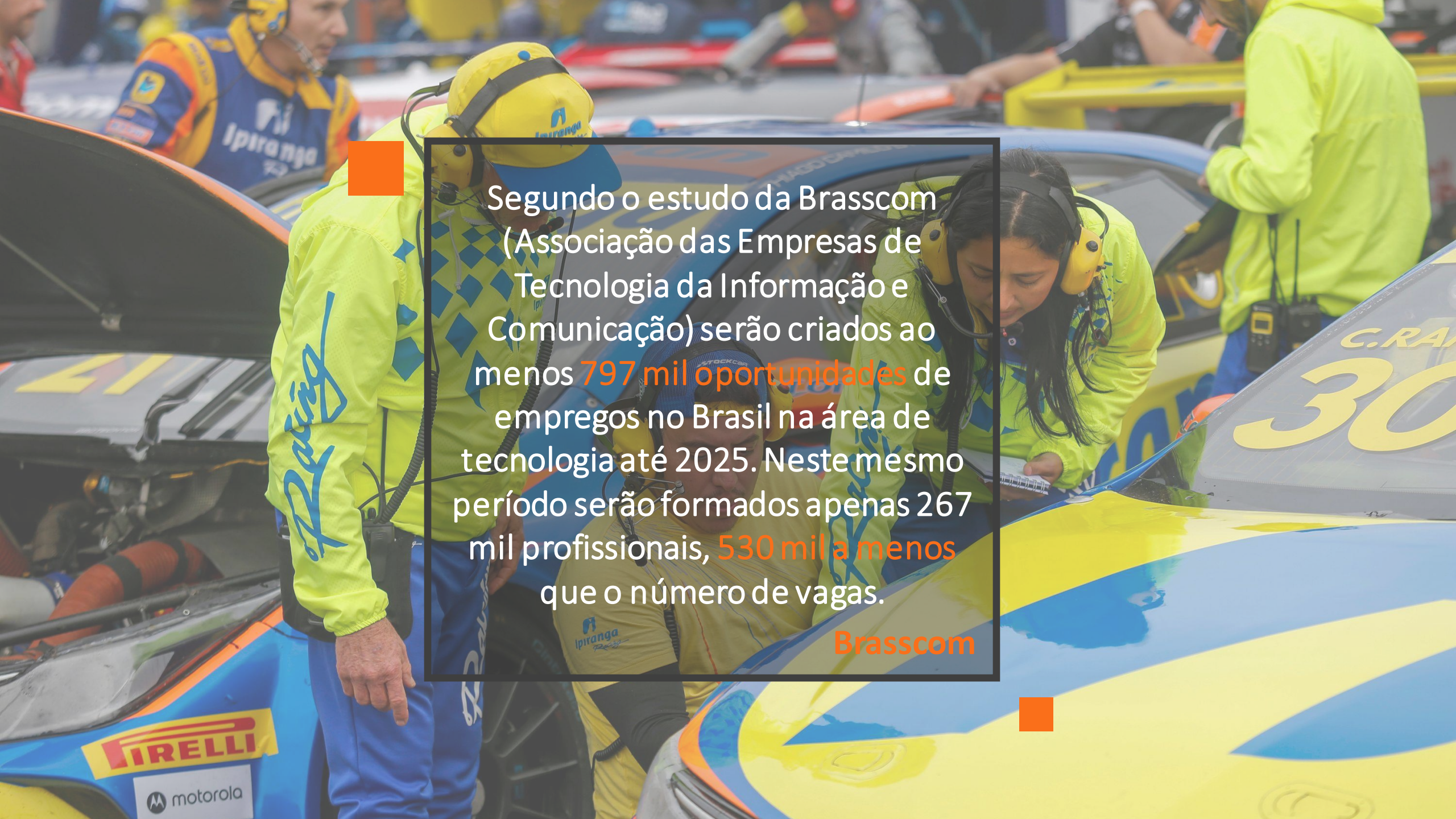
- Parte 1: Introdução ao Python
 - Linguagem de programação
 - Ambiente de desenvolvimento
 - Notebook
- Parte 2: Análise exploratória
 - Um projeto de Data Science
 - O que é um Dataframe
 - Visualização de dados



Agenda

- Parte 3: Análise bi-variada
 - Passo a passo de um projeto
 - Datasets
 - Tipos de dados
- Parte 4: Correlação de variáveis
 - Estatística
 - Vieses
 - Tipos de dados





Segundo o estudo da Brasscom (Associação das Empresas de Tecnologia da Informação e Comunicação) serão criados ao menos 797 mil oportunidades de empregos no Brasil na área de tecnologia até 2025. Neste mesmo período serão formados apenas 267 mil profissionais, 530 mil a menos que o número de vagas.

Brasscom



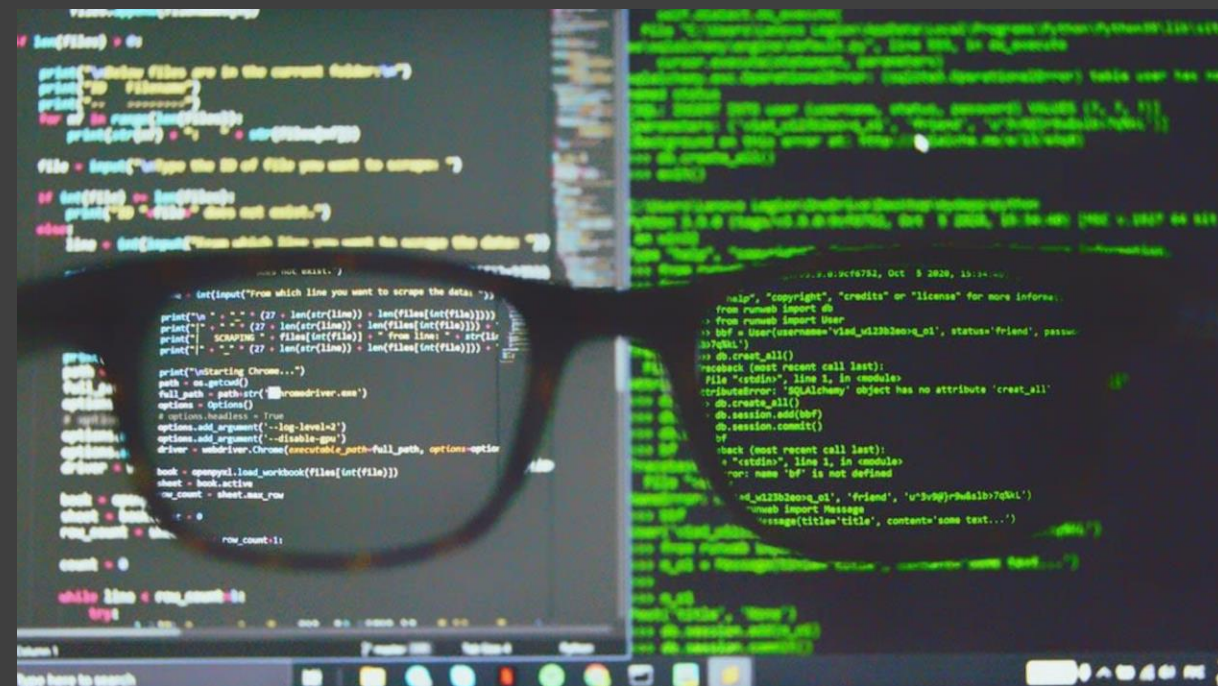
Introdução ao Python

Introdução ao Python



Linguagem de programação

- Conjunto de regras para implementação de um código fonte que pode ser compilado e transformado em um programa de computador.
- Linguagens de alto nível são escritas por programadores e posteriormente compiladas para linguagem de baixo nível que são interpretadas e executadas pelo microprocessador.
- Python é uma linguagem de alto nível



Introdução ao Python



Algoritmo

- É definido como uma sequência finita de ações executáveis.
- Suas ações são executadas ordenadamente.
- Na ciência da computação, é representado por um código em uma linguagem de programação e deve ter a finalidade de solucionar determinados problemas.

```
def train_model(train_data, val_data, model, lr=0.01, momentum=0.9, nesterov=False, n_epochs=30):  
    " Treinar o modelo por N épocas com os dados e os hiper parâmetros "  
    # Utilizando o otimizador Stochastic Gradient Descent  
    optimizer = torch.optim.SGD(model.parameters(), lr=lr, momentum=momentum, nesterov=nesterov)  
  
    for epoch in range(1, 101):  
        print("-----\nEpoch {}: \n".format(epoch))  
        # Treinamento  
        loss, acc, error, sochat = run_epoch(train_data, model.train(), optimizer)  
        print('Train loss: {:.6f} | Train accuracy: {:.6f}'.format(loss, acc))  
  
        # Validação  
        val_loss, val_acc, error, sochat = run_epoch(val_data, model.eval(), optimizer)  
        print('Val loss:  {:.6f} | Val accuracy:  {:.6f}'.format(val_loss, val_acc))  
  
        # Salvando o modelo  
        torch.save(model, 'mnist_model_fully_connected.pt')  
    return val_acc
```

Introdução ao Python



Ambiente de desenvolvimento

- Local onde iremos escrever as instruções que serão executadas pelo programa.
- Compila o programa e exibe seus resultados.
- Importar arquivos e bibliotecas.
- Possibilita execução linha por linha para a verificação da lógica e possíveis erros (breakpoints).



Introdução ao Python



Notebook

- Forma interativa de compartilhar documentos com códigos, gráficos, resultados, textos e outras funcionalidades.
- Permite a execução de um bloco de códigos.
- Os códigos podem ser alterados e executados a qualquer momento.
- O Google Colab é um exemplo de notebook, que será utilizado nesta aula.

+ Código + Texto

Imersão Motorsports para Mulheres

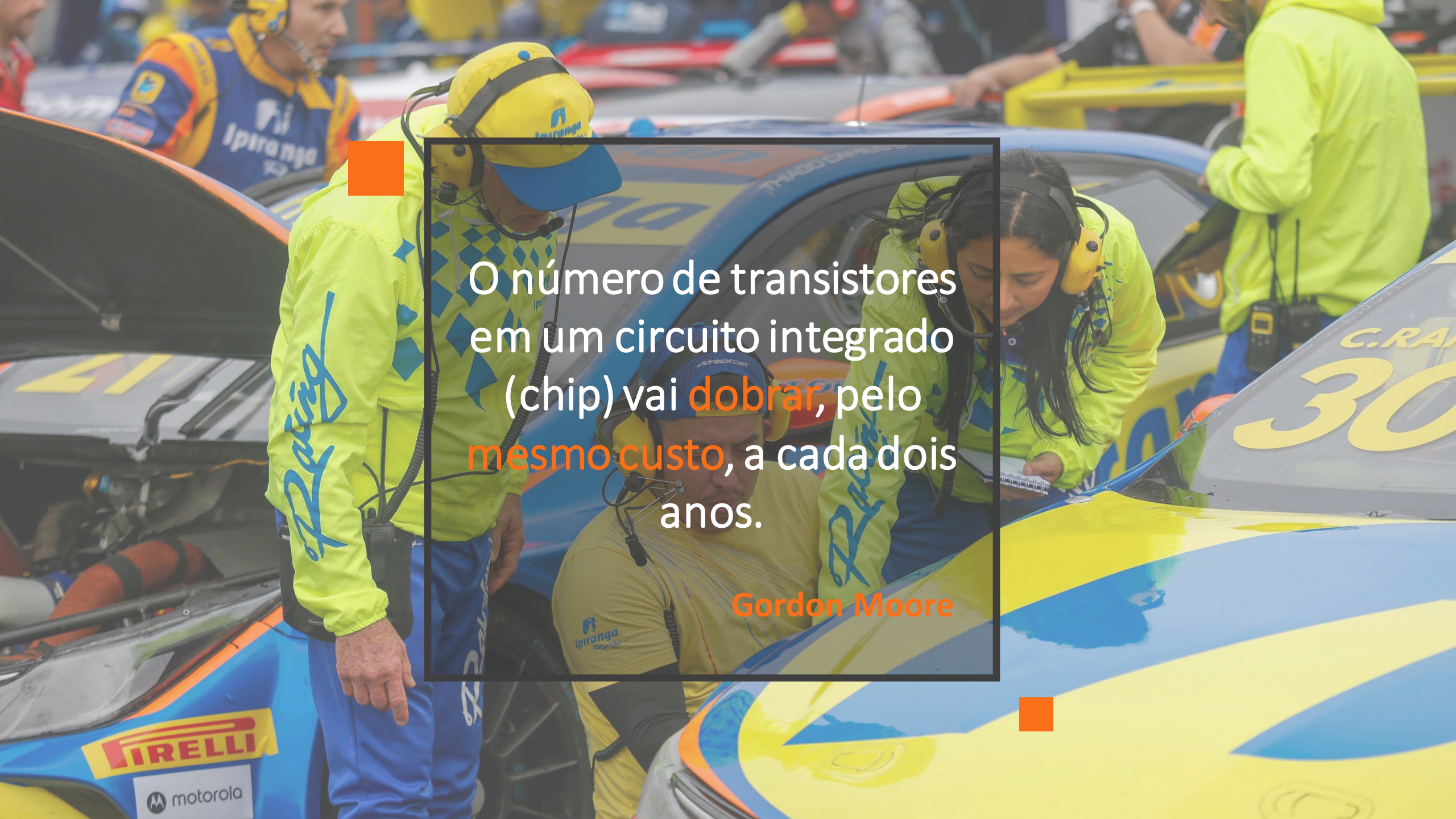
Ambiente de desenvolvimento - Google Colaboratory

Clique duas vezes (ou pressione "Enter") para editar

```
[ ] nome=input("Qual é o seu nome? ")
resposta=input("Você é apaixonada por Motorsports? ")
#print('\nOlá,', nome, '! É um prazer ter você como aluna na Imersão no Motorsports para Mulheres - AMattheis')
#print('\nOlá, ' + nome + '! É um prazer ter você como aluna na Imersão no Motorsports para Mulheres - AMattheis')
print(f'\nOlá, {nome}! É um prazer ter você como aluna na Imersão no Motorsports para Mulheres - AMattheis')
```

Qual é o seu nome? Sonia
Você é apaixonada por Motorsports? Sim

Olá, Sonia! É um prazer ter você como aluna na Imersão no Motorsports para Mulheres - AMattheis



O número de transistores
em um circuito integrado
(chip) vai **dobrar**, pelo
mesmo custo, a cada dois
anos.

Gordon Moore



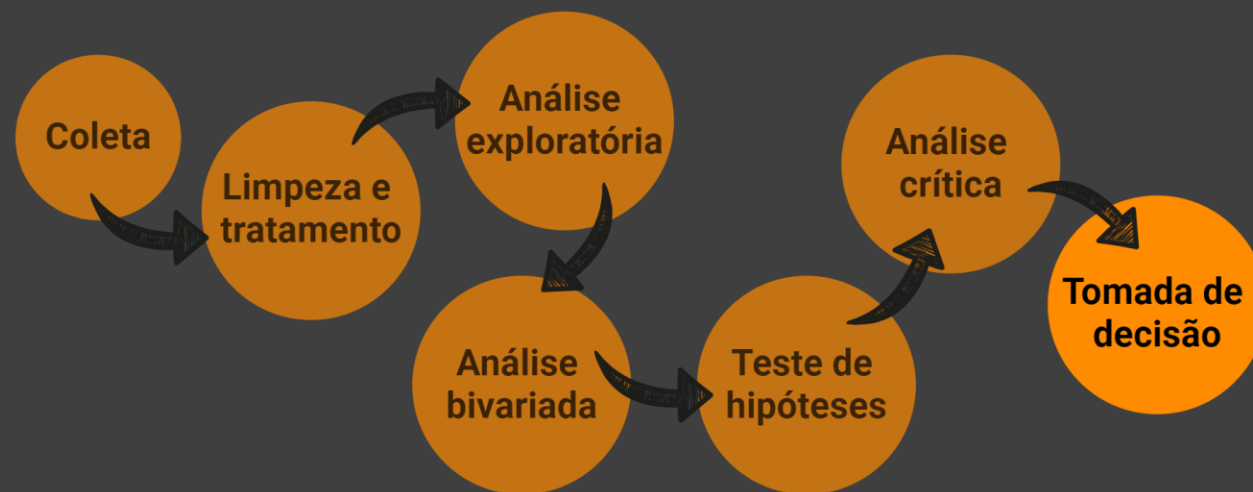
Análise exploratória

Análise exploratória



Um projeto de Data Science

- O Data Science é um meio e não um fim, é uma forma que temos de chegar as respostas que precisamos, o final sempre deve ser uma ação, uma tomada de decisão.
- Perguntas que devemos fazer:
 - **QUAL** é o problema que vamos resolver?
 - **PORQUE** precisamos resolver?
 - **ONDE** estão os dados que precisamos?
 - **COMO** vamos coletar esses dados?



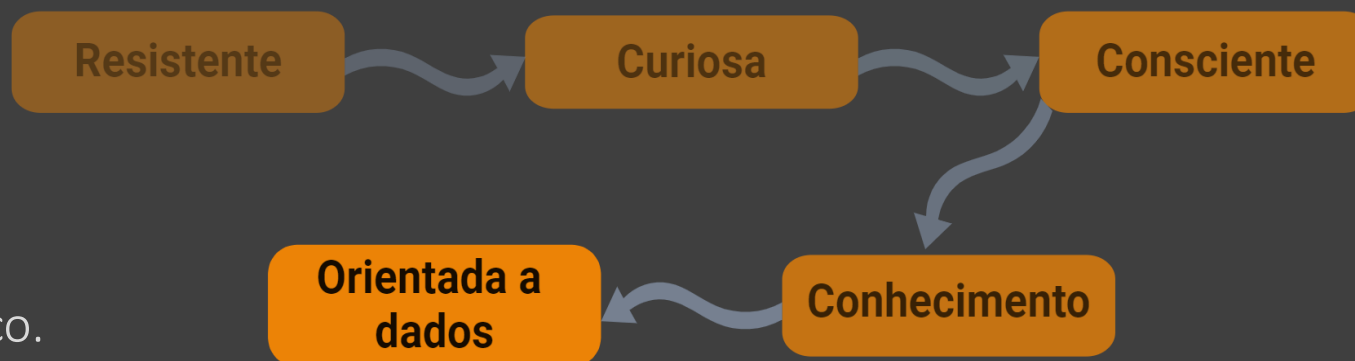
Análise exploratória



Cultura Data Driven

- Tomada de decisão orientada por dados, sendo eles a base de qualquer tomada de decisão.
- **Resistente:** "Sempre foi assim".
- **Curiosa:** Coleta dados, mas sem estrutura.
- **Consciente:** Analisa os dados.
- **Conhecimento:** Dados tem valor estratégico.
- **Orientada a dados:** Coleta de dados estruturada, uso de tecnologia e IA, tomada de decisão baseada em dados.

Evolução das empresas



Análise exploratória



O que é um Dataframe

- É uma estrutura tabular, formada por linhas e colunas.
- Cada coluna possui um header (cabeçalho)
- Toda linha possui um index
- Em Python a biblioteca Pandas possui a funcionalidade Data frame, muito utilizada para a manipulação e análise de dados.

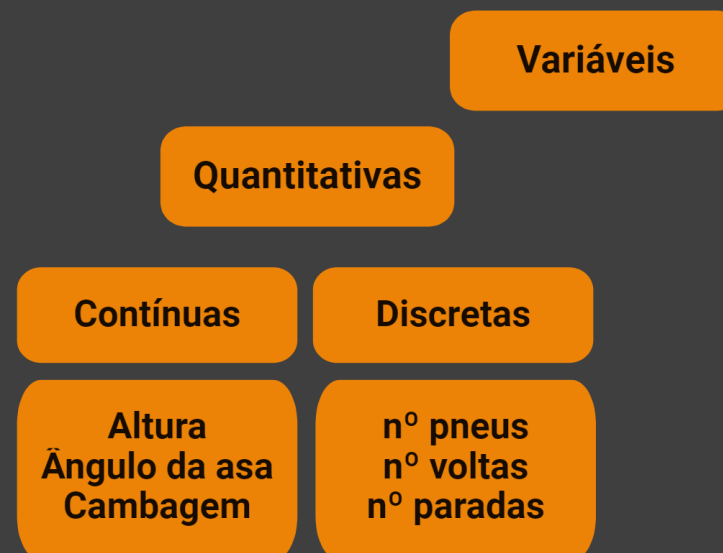
index					header

Análise exploratória



Tipos de variáveis

- Quantitativas: variáveis numéricas
- Tipos
 - Discreta:
 - Número inteiro
 - Ex: Quantidade de pneus
 - Contínua:
 - Números em escala contínua (casa decimal)
 - Ex: Altura dianteira do carro



Análise exploratória



Tipos de variáveis

- Qualitativas: descrevem uma qualidade/atributo
- Tipos
 - Nominal:
 - Não existe uma ordem
 - Ex: Corola ou Cruze
 - Ordinal:
 - Existe uma ordem
 - Ex: piloto preenchendo no debriefing o nível de OS na curva: pouco, médio ou alto.



Análise exploratória



Vizualização de dados

- Através dos gráficos vamos transmitir um resultado, fazer comparações, observar tendências, levantar hipóteses...
- Tipos de gráficos:
 - Gráfico de barras
 - Gráfico de linhas
 - Histograma
 - Boxplot
 - Scatter Plot (dispersão)
 - Gráfico de Pizza (não recomendado)

Gráfico de barras

Comparar categorias em relação a uma variável.

Setor 1 dos carros

Velocidade das marcas

Gráfico de dispersão

Comparar a relação entre duas ou mais variáveis.

Velocidade x gap

Tempo de volta x nº volta



Gestalt

As leis básicas da Gestalt são:

- Semelhança
- Proximidade
- Continuidade
- Pregnância
- Fechamento
- Unidade
- Unificação
- Segregação



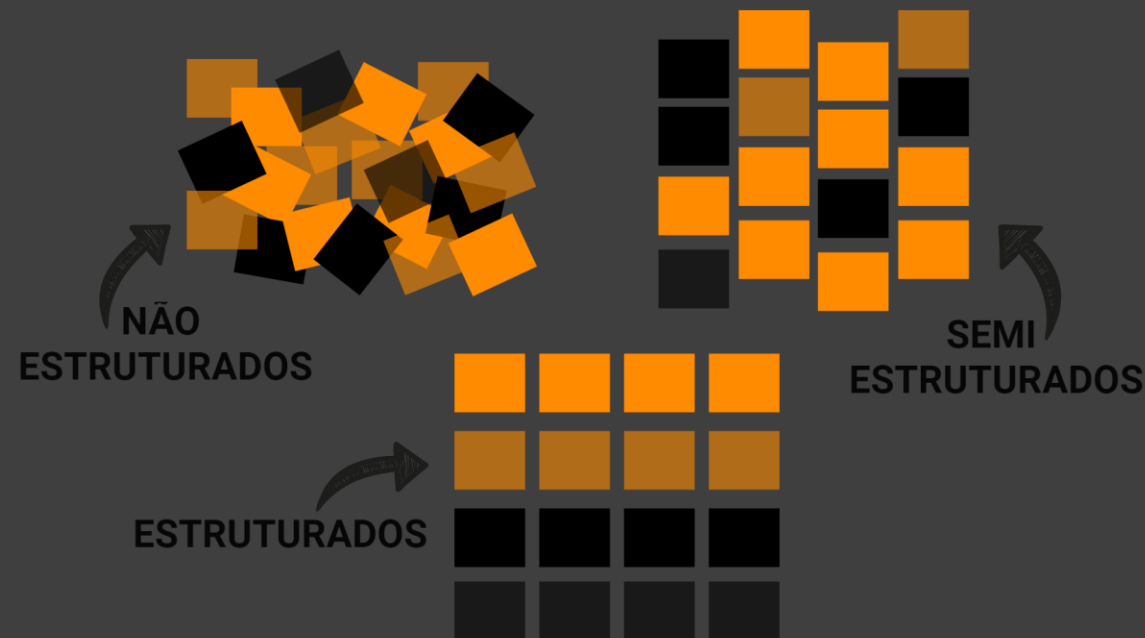
Análise bivariada

Análise bivariada



Tipos de dados

- Dados estruturados
 - Organizados em linhas e colunas, em forma de uma matriz
 - Exemplo: tabela no Excel, dataframe no Pandas
- Dados semi-estruturados
 - Possui uma estrutura/organização, mas fora do padrão de tabelas/matrizes com linhas e colunas
 - Exemplo: arquivos .json
- Dados não estruturados
 - Não possuem estrutura definida
 - Exemplo: imagens, áudios e vídeos



Análise bivariada



Coleta de dados

- Antes de qualquer análise precisamos ter confiabilidade com os dados que estamos analisando
- Perguntas que devemos fazer:
 - Este sensor foi calibrado corretamente?
 - O sensor está na posição certa?
 - A taxa de amostragem é adequada?
 - A fonte de dados externa é confiável?
 - Os dados possuem algum viés?

TEOREMA DE NYQUIST

A taxa de amostragem (f_a) deve ser maior que o dobro da componente de maior frequência que você quer analisar no sinal medido (f_n)

$$f_a > 2 * f_n$$

Análise bivariada



Datasets

- Fonte de dados estruturada relacionada a um tema específico onde podemos aplicar técnicas de análise de dados.
- Fontes de dados que podem gerar datasets:
 - Software de análise de dados
 - Sistema de cronometragem
 - Planilhas de setup
 - Tabelas de controle de peças
 - Dados históricos de tempos e resultados

CHRONON

STOCK PRO SERIES - 2022

INÍCIO » RESULTADOS » STOCK » STOCK PRO SERIES - 2022

São Paulo Goiânia Rio de Janeiro VeloCitta Velopark São Paulo Velocitta Sta Cruz Goiânia São Paulo

#1 SÃO PAULO - 13 DE FEVEREIRO

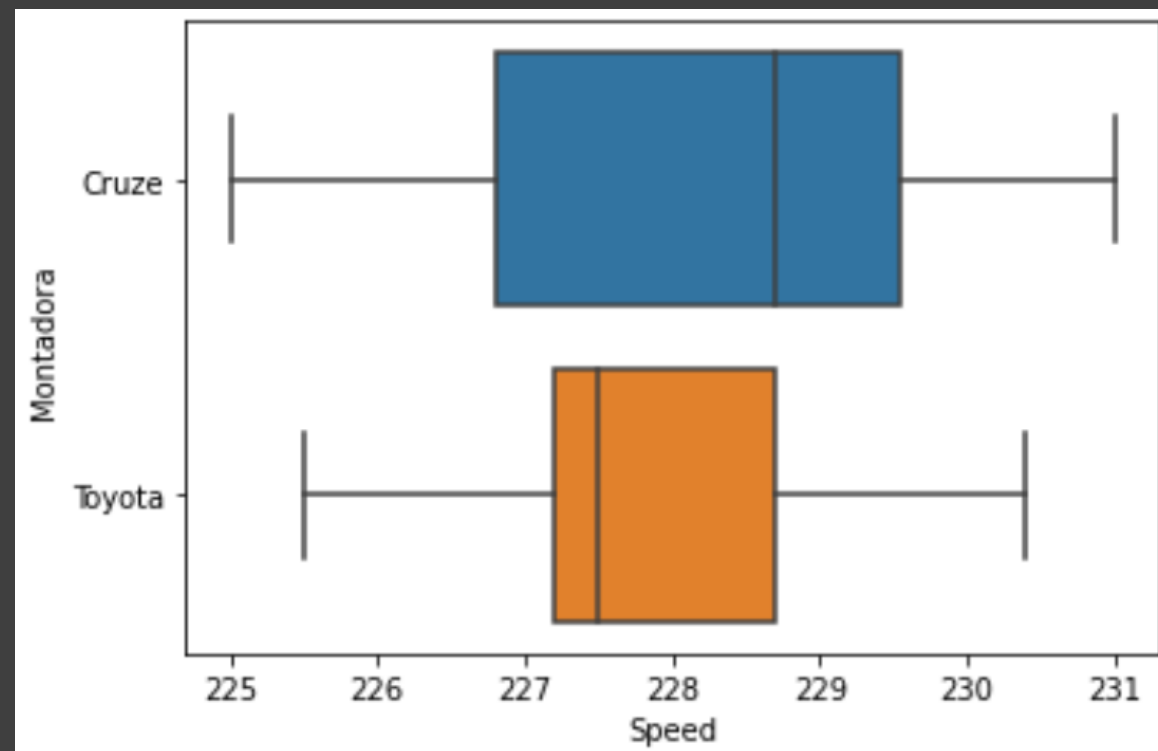
	RESULT	SECTOR	SECTLAP	LAPS	SPEED	MAP	GRID	TRACK	PIT	ALL FILES
Shake Down	PDF	PDF	-	PDF CSV	CSV	-	-	-	-	-
TR	PDF	PDF	-	PDF CSV	CSV	-	-	-	-	-
T1	PDF	PDF	-	PDF CSV	CSV	-	-	-	-	-
T2	PDF	PDF	-	PDF CSV	CSV	-	-	-	-	-
T3	PDF	PDF	-	PDF CSV	CSV	-	-	-	-	-

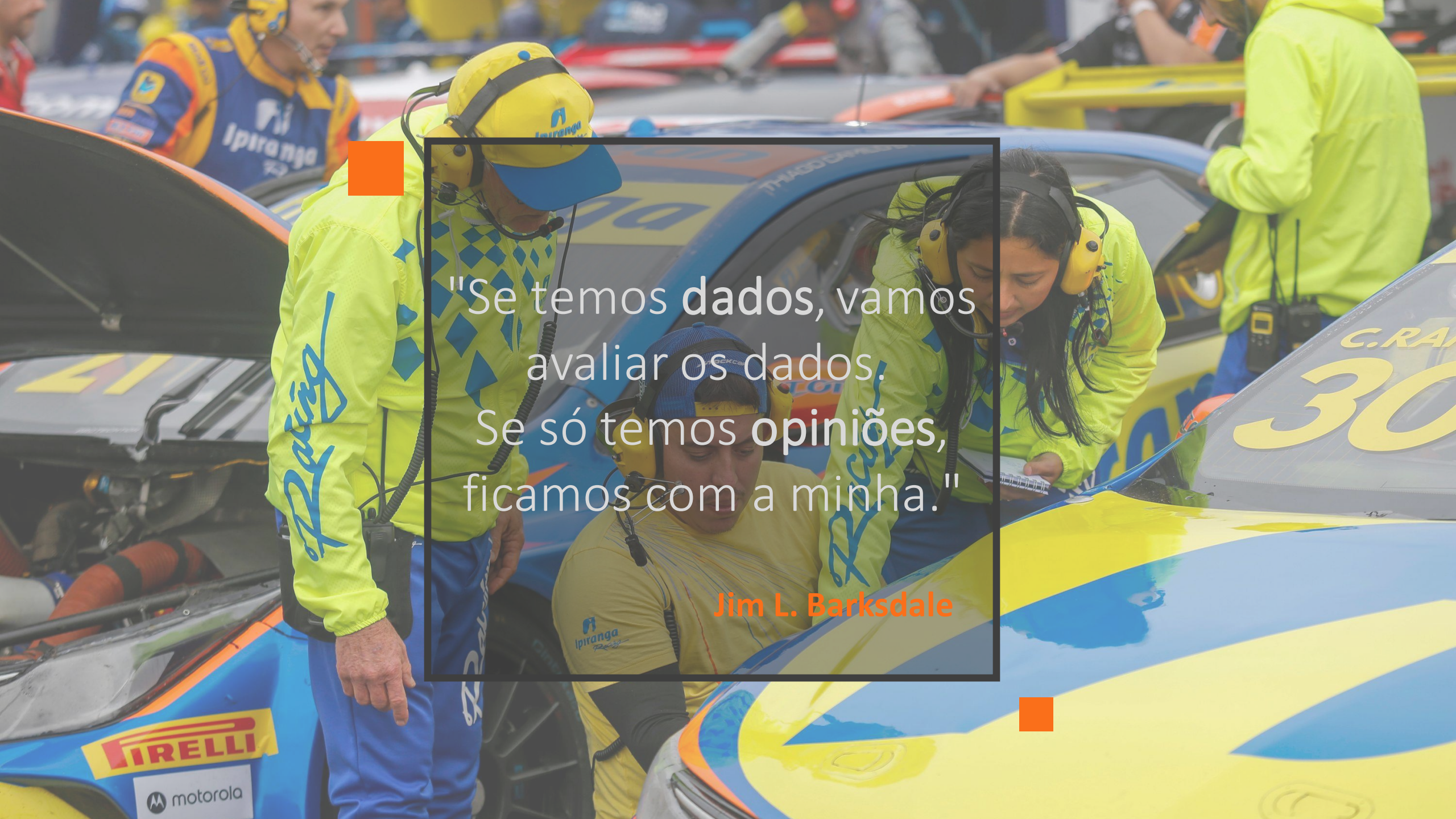
Análise bivariada



Gráficos de correlação

- Variáveis numéricas
 - Gráficos de dispersão (Scatter plot) são ótimos para encontrarmos tendências.
 - Exemplo: Velocidade x Gap.
- Variáveis categóricas
 - Gráficos de barras agrupadas.
 - Podemos observar a maior incidência de uma variável categórica em relação a outra.
- Variável numérica com variável categórica
 - Utilizar o gráfico boxplot e observar as diferenças dos percentis, mediana e outliers.





"Se temos dados, vamos
avaliar os dados.
Se só temos opiniões,
ficamos com a minha."

Jim L. Barksdale



Correlação de variáveis

Correlação de variáveis



Teste de hipóteses

- Uma hipótese é baseada em um KPI
 - "A nova carenagem aumenta em 5% o **downforce** dianteiro."
 - "O modelo antigo da asa era 4% menos eficiente que o atual."
 - "Este pneu melhora em 3% o **tempo de volta**."
- O teste de hipótese verifica se existe relação estatística entre os dados analisados.

2 variáveis
numéricas

Pearson
Spearman

2 variáveis
categóricas

Qui-quadrado

2 grupos
independentes

Teste-t
independente

+ de 2 grupos
independentes

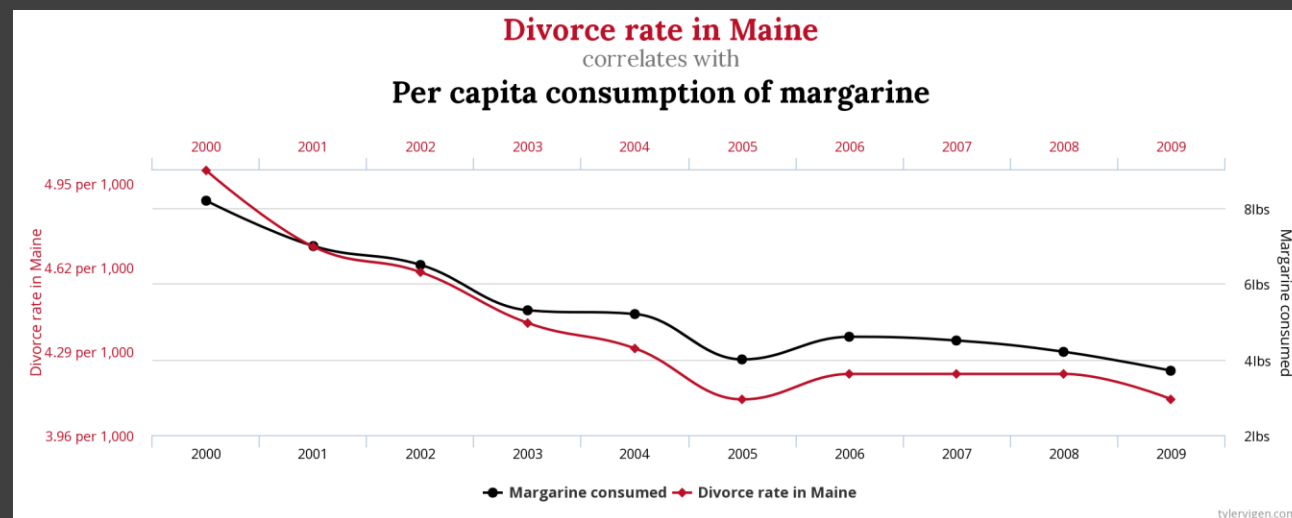
ANOVA

Análise exploratória



Correlação x Casualidade

- Correlação **NÃO** É casualidade.
- Dizemos que existe uma **correlação** entre duas variáveis se suas variações em função do tempo representam algum grau de semelhança. Podemos classificar correlações do tipo forte, fraca, nula, exponencial, linear ou em forma de U, por exemplo.
- Dizemos que existe uma **casualidade** entre duas variáveis quando a variação de uma variável causa a variação da outra.



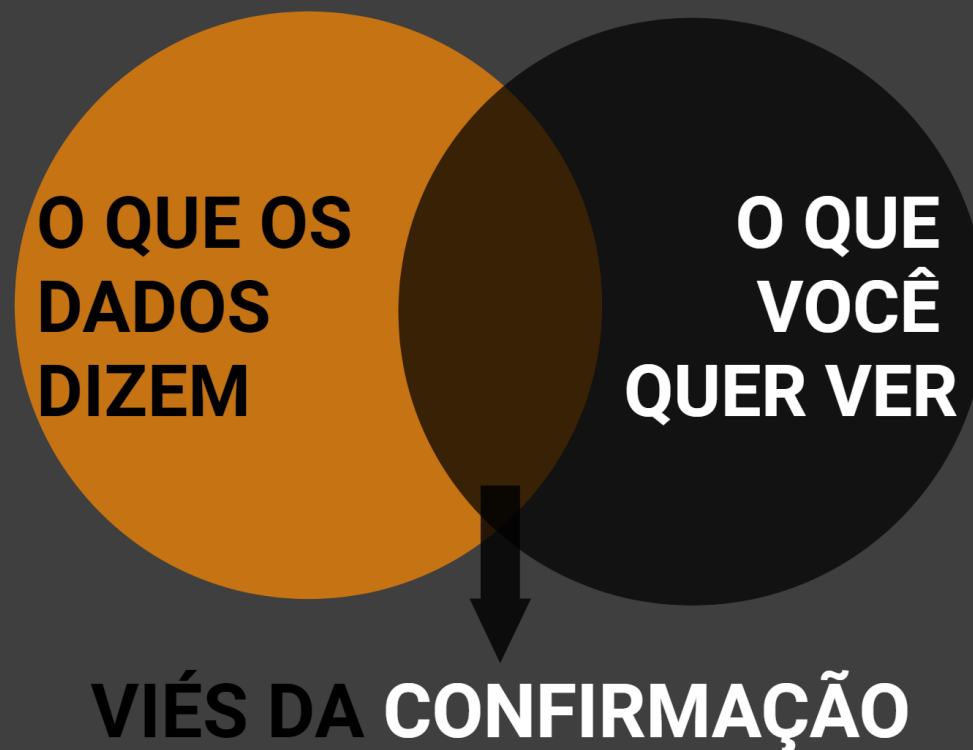
<https://tylervigen.com/spurious-correlations>

Análise exploratória



Vieses

- Eles podem distorcer e influenciar a sua percepção e análise dos dados.
- Tipos de vieses:
 - Amostragem
 - Quando chegamos a conclusões generalistas partindo de uma amostra pequena de dados, que não representa o todo.
 - Entrevistador
 - Comum quando o engenheiro faz um debriefing com o piloto e faz perguntas sem imparcialidade, influenciando a resposta do piloto.
 - Confirmação
 - Você faz alguma mudança no setup do carro esperando uma determinada resposta e quando vai analisar os dados procura confirmar o que você espera.



BÔNUS!

Inteligência Artificial



Computação tradicional

O que conhecemos até agora

- Forcemos os dados de entrada e as regras do algoritmo.
- A saída do nosso modelo é uma resposta baseada nas regras que definimos
- Regras do problema são definidas pelo programador e são fixas no algoritmo.



Aprendizado de máquina



O que podemos aprender

- Forcemos os dados de entrada e os dados de resposta para o computador
- A saída do nosso modelo é são as regras que correlacionam os dados de entrada e resposta
- Quem define as regras do problema é a maquina (computador), que vai aprendendo durante o processo de treinamento e ajustando seus parâmetros.

**Dados
de entrada**

**Dados
de resposta**

Computador

Regras

Comparação



Programação Tracional



Aprendizado de Máquina

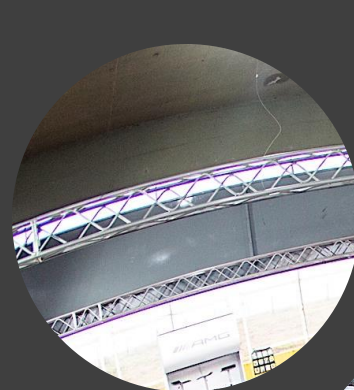




Como podemos aplicar aprendizado de máquina em Motorsports?

Resumo

- Aprendemos sobre
 - Linguagem de programação em Python
 - Projeto de ciência de dados
 - Análise de dados no Colab
 - Bibliotecas Pandas, Numpy, Matplotlib e Seaborn
 - Criar gráficos e analisar variáveis com Python
 - Fazer correlação de variáveis e levantar hipóteses
 - O que é aprendizado de máquina





16 de abril de 2000.

Todas as 31 franquias
da NFL tiveram a
oportunidade de escolher
Tom Brad, mas coube aos
Patriots selecionarem-no,
com a 199ª escolha geral.

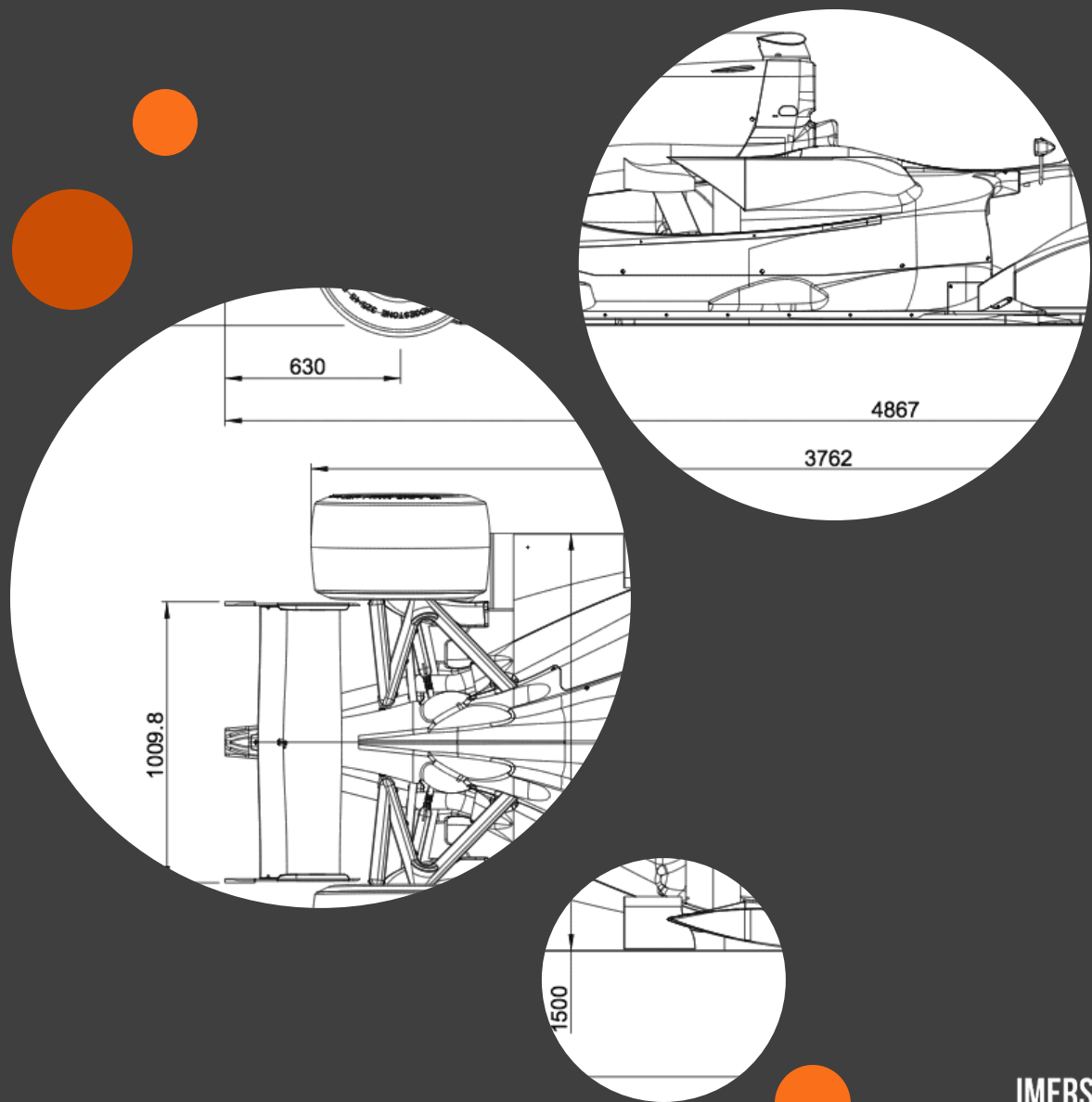
Dúvidas?

Muito obrigado a todas as mulheres que assistiram essa aula, desejo todo sucesso na jornada de cada uma e dizer que podem contar comigo nessa caminhada.

Agradecimento especial a Rachel Loh, demais palestrantes, equipe da produção e toda equipe AMattheis pela oportunidade de estar aqui compartilhando com vocês.

Fábio Mori

fsmmori@gmail.com





Fábio Mori

Programação Aplicada

IMERSÃO NO
MOTORSPORT
PARA MULHERES

A MATTHEIS
MOTORSPORT