

Análise dos Dados do Bot no Carnaval de Belo Horizonte

Processo Seletivo para Vaga de Engenharia de Dados
Candidato: Renan Santos Mendes

June 1, 2018



Sumário

- 1 Pergunta 1
- 2 Pergunta 2
- 3 Pergunta 3
- 4 Pergunta 4
- 5 Pergunta 5
- 6 Novas hipóteses
- 7 Hipóteses não validadas

Pergunta 1

Qual a quantidade de mensagens dos usuários?



Resposta 1

Filtrando as mensagens, temos que há um total de 11022 mensagens enviadas pelos usuários.



Pergunta 2

Quantos são os usuários únicos do Bot?



Pergunta 3

Qual a quantidade de mensagens trafegadas?



Resposta 3

Somando as mensagens enviadas pelos usuários e pelo bot, temos um total de 33766 mensagens.



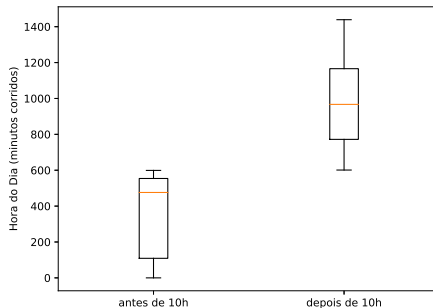
Pergunta 4

Apresente, da forma como achar melhor, as interações que os foliões tiveram com o Bot.



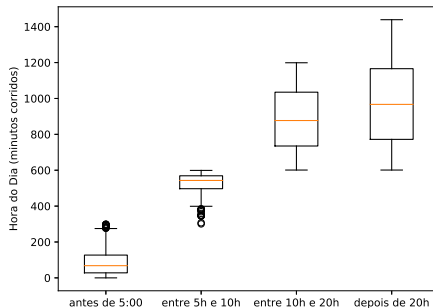
Resposta 4

- Boxplot para os dois intervalos em análise no carnaval.



Resposta 4

- Boxplot para os novos intervalos em análise no carnaval.



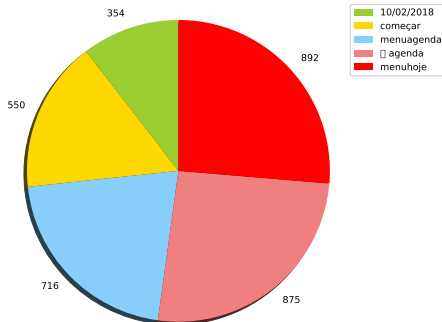
Pergunta 5

Quais botões foram os que tiveram maior interação?



Resposta 5

- Quantidade de vezes que os 5 botões mais utilizados foram pressionados



- Análise foi feita com base nos botões apresentados no Fluxo Figma;
- O botão HOJE, denominado nos dados como MenuHoje, foi o mais utilizado, sendo apertado 892 vezes;
- Isso mostra que os usuários se interessaram pela programação do dia apesar da grande utilização da agenda (875 vezes).

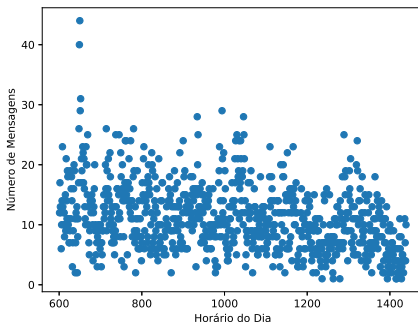
Novas Hipóteses

Novas Hipóteses a serem inferidas sobre os dados



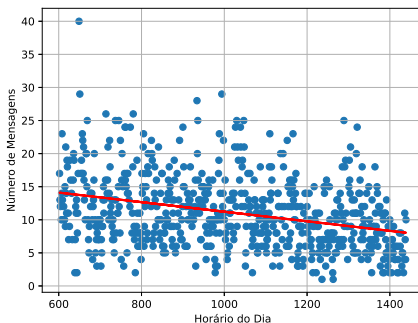
Hipótese 3 - É possível descrever o comportamento da utilização do bot usando uma regressão linear?

- Para termos uma maior noção do que acontece com os dados, vamos fazer um scatterplot para ver a dispersão.



Hipótese 3 - É possível descrever o comportamento da utilização do bot usando uma regressão linear?

- Os dados foram separados em teste e treino;
- Os dados de treino foram utilizados para gerar a seguinte reta:



Hipótese 3 - É possível descrever o comportamento da utilização do bot usando uma regressão linear?

- Não houve um bom ajuste do modelo;
- Uma alternativa seria alterar os dados, uma vez que estão apresentados em minutos, seria viável agregar eles por faixa horária (5 em 5 minutos, 10 em 10 minutos, dentre outras);
- Uma segunda alternativa seria em abrir os dados para cada dia, considerando os dias mais utilizados do bot, assim evitando influência de dias em que o bot foi pouco utilizado.

Hipóteses não validadas

Outras hipóteses formuladas para serem testadas futuramente



Hipóteses não validadas

- 1 O bot atendeu a necessidade do usuário - pensar em alguma forma de feedback ao fechar o bot
- 2 Haveria como localizar de onde o usuário utilizou o bot? IP?
- 3 Haveria diferença nos resultados obtidos ao se comparar cada dia do carnaval? (pensando numa análise micro dos dados).
- 4 Seria possível fazer uma análise de sentimento da interação do usuário com o bot? (para entender o comportamento do usuário).

Obrigado pela atenção!

