



Turing Academy - 2023 - Entregável I

Análise/Limpeza de Dados

Chegou o momento! Os entregáveis que você, trainee, deverá realizar consistem em um conjunto de 3 exercícios que avaliarão o seu domínio nos tópicos estudados durante o Turing Academy até o momento, dentre os quais:

- Manipulação de DataFrame com pandas;
- Visualização;
- Limpeza;
- Análise;
- Pré-Processamento de dados;
- Básico de Machine Learning;

Para não sobrecarregá-los, soltaremos um exercício por vez (1 por semana). Recomendamos fortemente que vocês tentem fazer os exercícios à medida que eles saiam (não deixem acumular!!!).

Para a entrega dos exercícios, pediremos para que vocês criem um repositório no GitHub (ele deverá ser público) e adicionem o link para ele no espaço apropriado dessa planilha: <https://docs.google.com/spreadsheets>. Submetam a esse repositório 1 arquivo no formato .ipynp para cada entregável. O prazo para entrega de todos os exercícios é até dia 09/07 (não necessariamente eles precisam ser entregues na ordem).

Para a realização dos exercícios, não se limitem aos conhecimentos vistos durante as aulas ou ao material do TA - incentivamos que busquem conhecimento em outras fontes, como no DataCamp ou no Youtube. Não hesitem em chamar os seus mentores ou qualquer um dos professores do TA para tirar dúvidas. Estaremos sempre à disposição! Bom trabalho!

EX1 - Um Experimento Esquisito

O excêntrico milionário John Turing decidiu conduzir um experimento inusual. Para isso, ele contratou como assistentes um grupo de 6 jovens universitários que, durante um mês inteiro, deveriam medir e registrar a temperatura de um lugar familiar a todos eles, a Praça do Relógio da USP. John solicitou que as medições fossem feitas 4 vezes por hora, 24 horas por dia, todos os dias durante esse mês. Os universitários foram alocados em um rígido cronograma, de forma que sempre houvesse um responsável por realizar as medições em um determinado horário. Embora o experimento fosse muito esquisito e à primeira vista sem propósito, nenhum dos universitários questionou, afinal estava sendo muito bem pagos para realizar a tarefa.



O grupo foi orientado a registrar a temperatura medida com precisão de 2 casas decimais e, caso esquecessem ou por algum motivo não pudessem realizar a medição em um determinado horário, eles deveriam deixar a temperatura para esse horário em branco. A seguir consta os dados coletados pelos estudantes ao longo do período de um mês: [dataset1.csv](#).

- Podemos notar que a coluna de data está formatada incorretamente. Utilize funções do **pandas** para convertê-la para o formato **datetime**.
- Esboce um gráfico da temperatura em função do tempo. O que você percebe analisando esse gráfico?
- John identificou que os dados enviados por um dos estudantes apresentou um comportamento suspeito. Você consegue identificá-lo?

- d) O dataset apresenta diversos valores ausentes de temperatura. Determine um método adequado para lidar com esses dados. (Dica: considere implementar uma solução de imputação com base em séries temporais.)
- e) Qual foi a quarta-feira mais fria do mês? Isto é, com a menor temperatura mínima? Qual foi essa temperatura?
- f) Por fim, construa um gráfico como o abaixo, para as temperaturas máximas e mínimas ao longo do mês.

