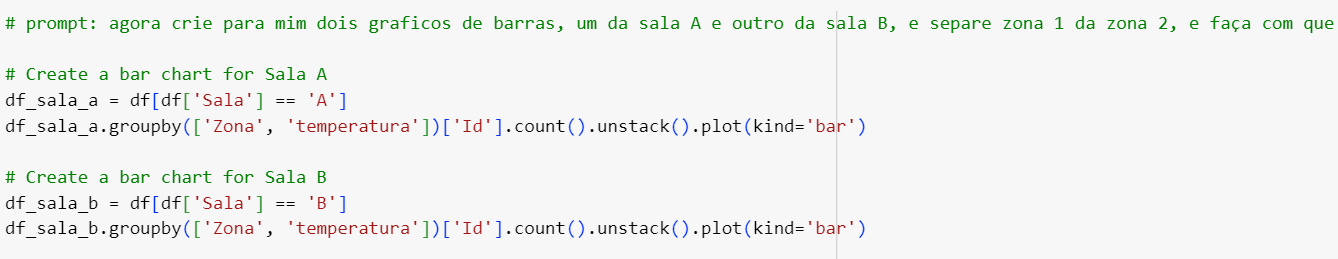
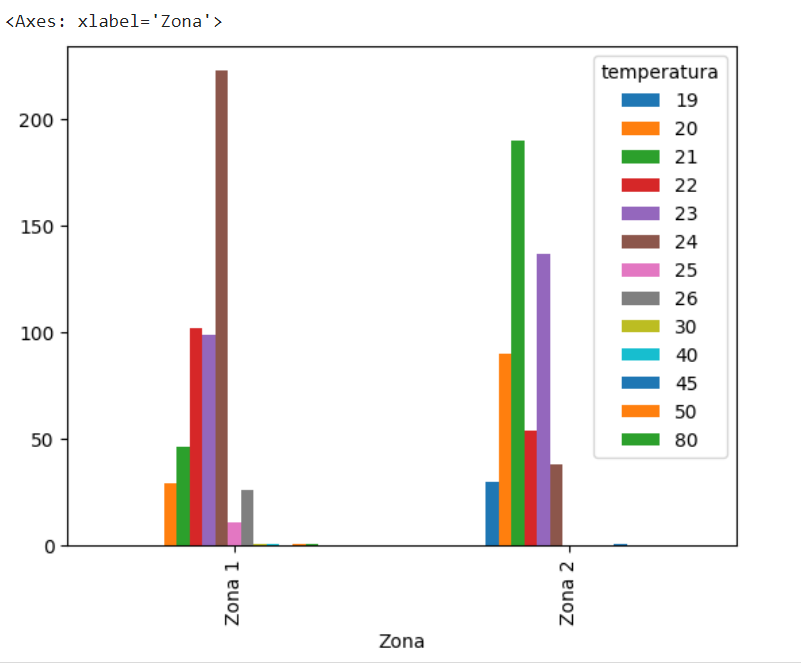
**ETAPA 02 – CIÊNCIA DE DADOS**

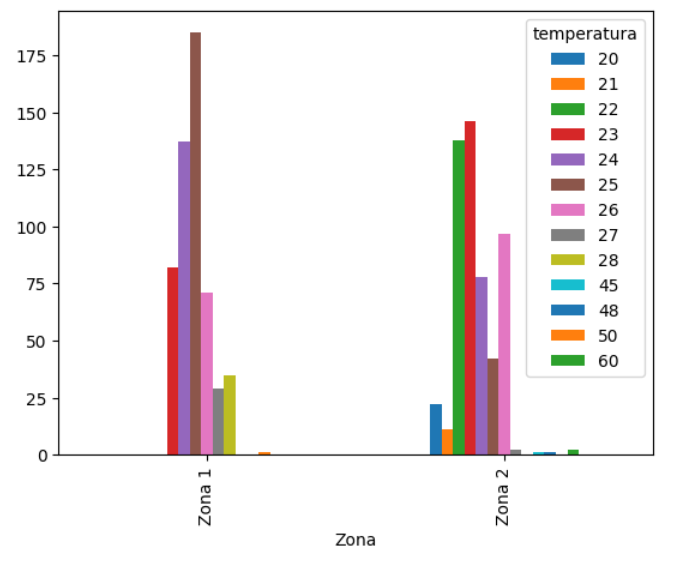
**1 - Formulação das hipóteses**:

 Para a formulação das hipóteses para descobrir se realmente há uma diferença significativa entre a sala A e B na questão da temperatura. Primeiro teremos que separar esses dados para ter mais precisão enquanto a quantidade de pessoas que correspondem a uma certa temperatura e assim separar elas em sala A e B, para isso utilizamos o Google Colab, onde pedimos para que ele por meio de IA generativa criasse um código em phyton que fizesse em forma de gráfico de barras essas divisão mais precisa dos dados que temos:

**SALA A:**

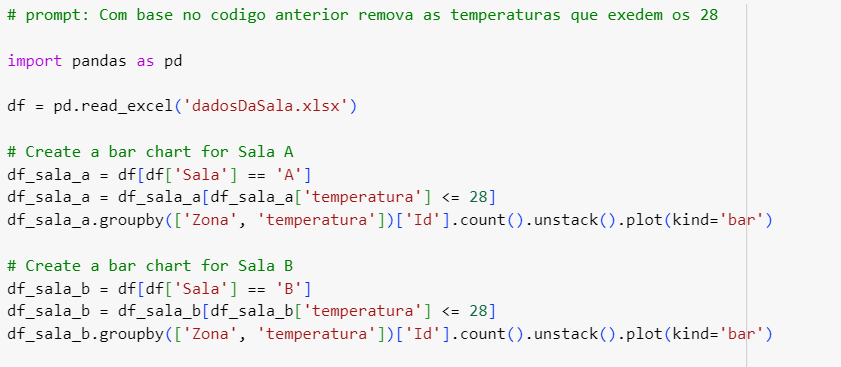


**SALA B:**

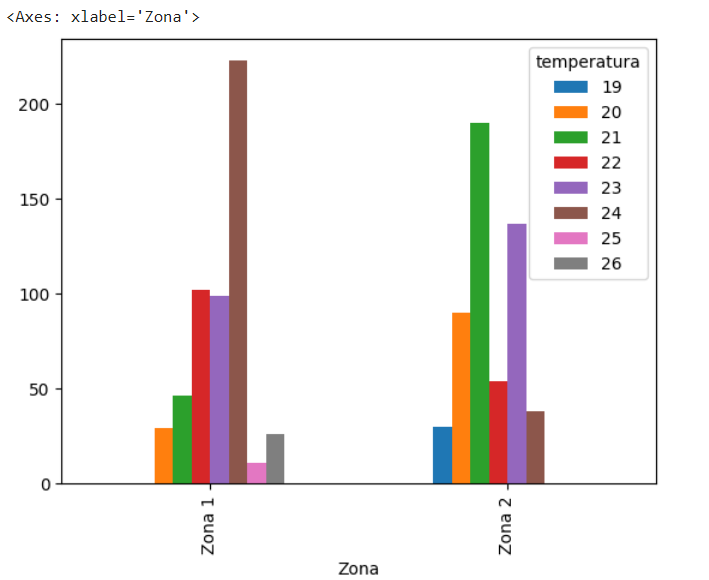


Ele gerou um código em phyton onde ele separa em duas tabelas, a tabela A e B, nelas podemos ver que as barras apresentadas mostram as temperaturas que foram apresentadas em cada sala e nelas mostra as variedades de cores para mostrar tais. Pode se notar também que ele criou no eixo Y a quantidade total de pessoas utilizando o Id delas que condiziam com a mesma temperatura, e já no eixo X ele separou por zonas na sala que são zona 1 e 2.

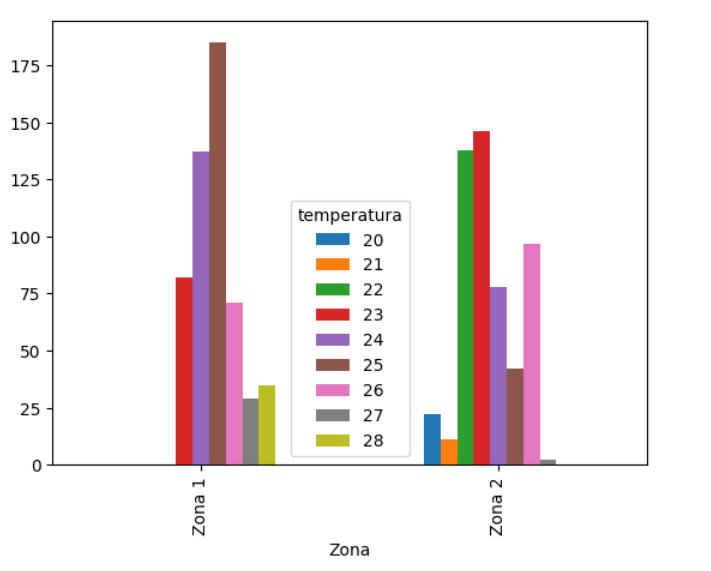
**1.2 Pré-processamento dos dados (verificar inconsistências):**

Pode se notar com base nas imagens dos gráficos anteriores que há uma inconsistência de dados enquanto a temperatura, porque ela ultrapassa o limite de um valor normal que seria entorno de 28°C graus, enquanto se mantermos a consistência pode se notar que de um minuto para outro a temperatura de 20 sobe para 50, e com base nisso tratamos essa inconsistência de dados e refizemos esse gráfico tirando os Id que continham tal inconsistência em suas temperaturas, e fatoramos uma parte do código:  
  


**SALA A:**

****

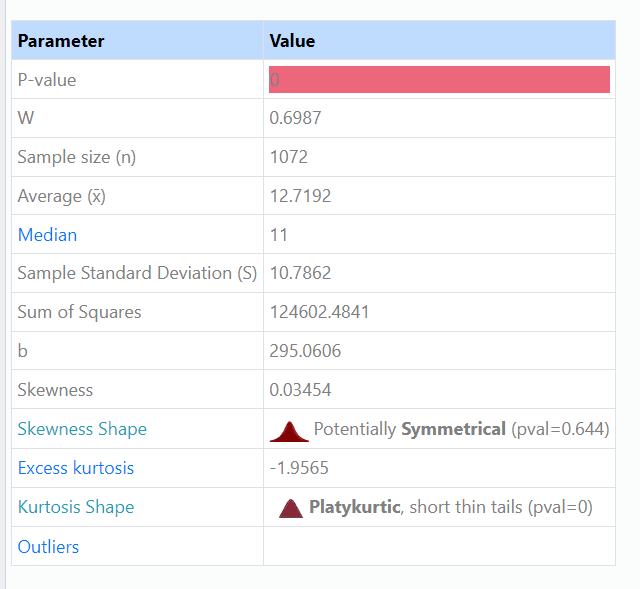
**SALA B:**



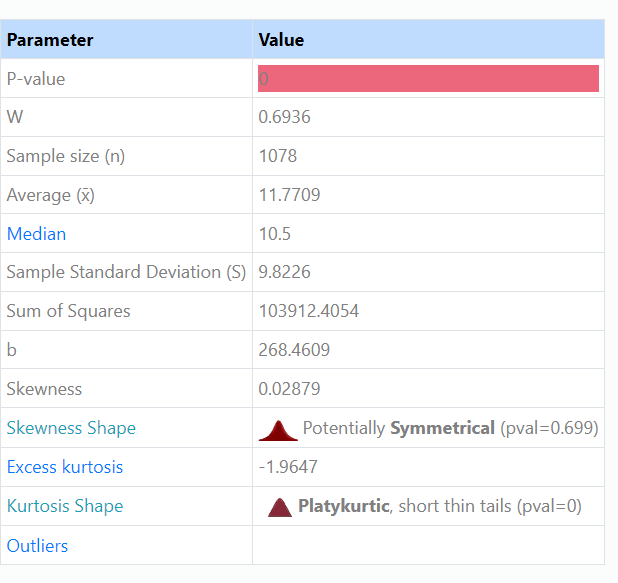
**2.0 Verificação dos parâmetros (tipos de dados, tipo de análise que queremos conduzir, normalidade):**

Para sabermos a normalidade de ambos os grupos tanto sala A e b, iremos fazer a análise de normalidade de cada sala com base em suas zonas e temperaturas, veja o resultado, utilizamos o modelo shapiro wilk para fazer essa normalidade:

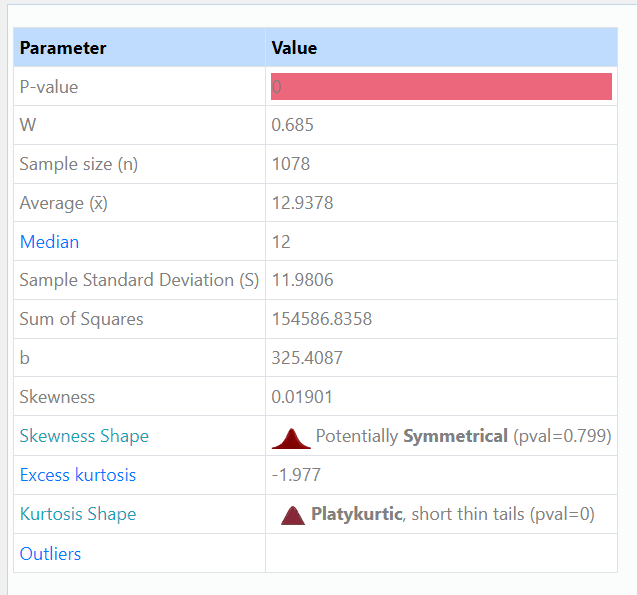
**SALA A - ZONA 1:**

****

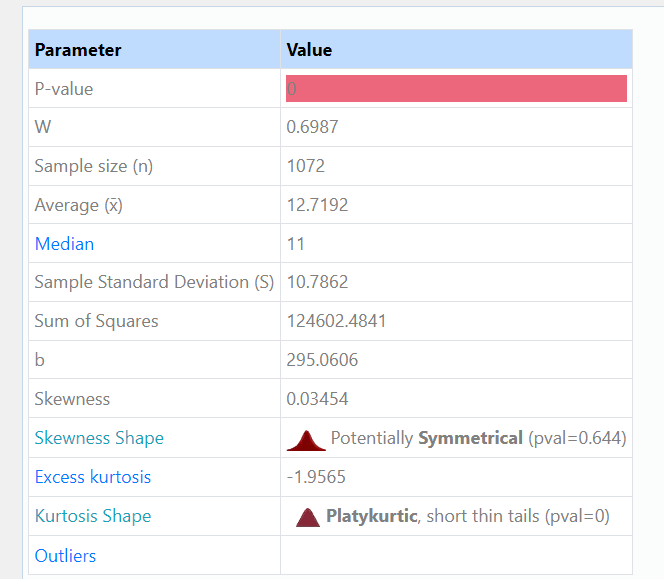
**SALA B – ZONA 2:**

****

**SALA B – ZONA 1:**

****

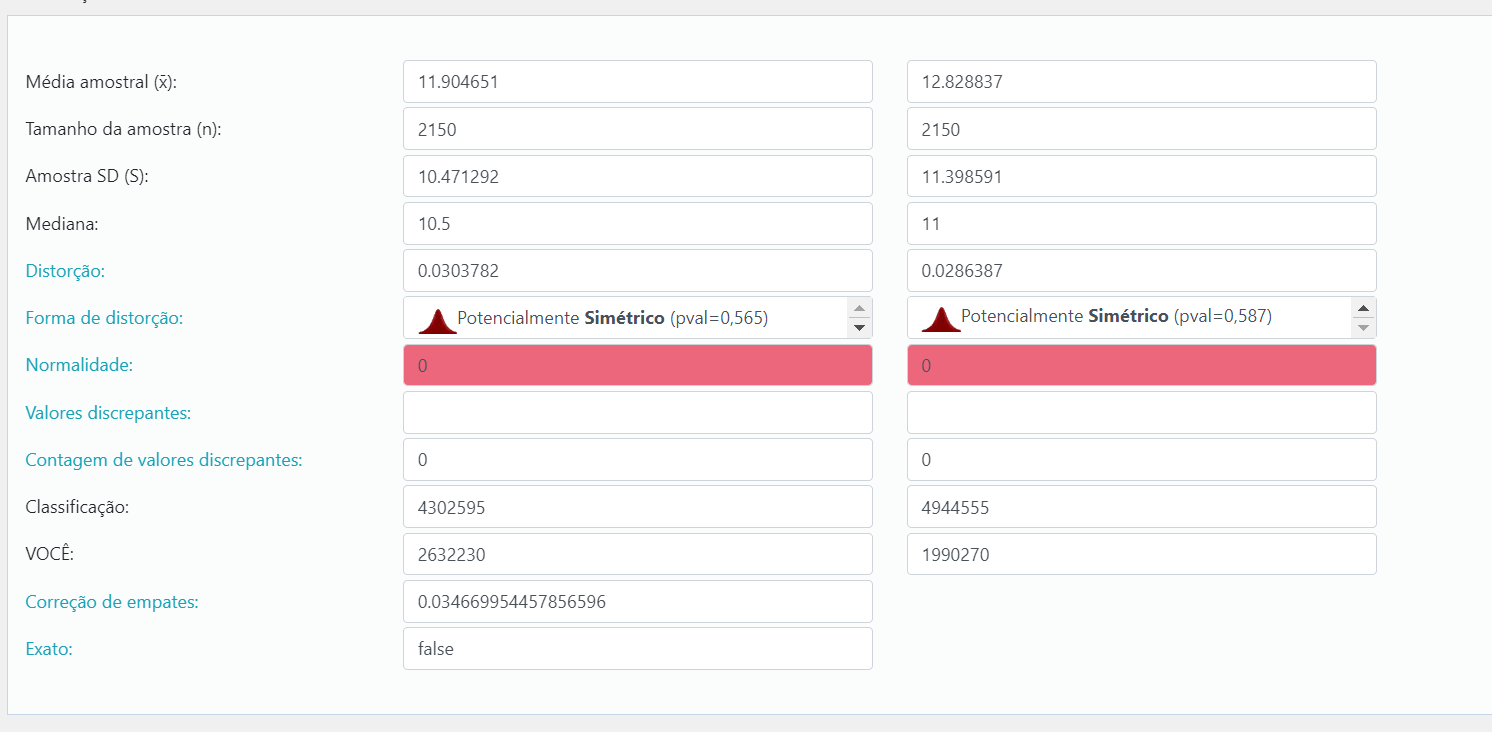
**SALA B – ZONA 2:**

****

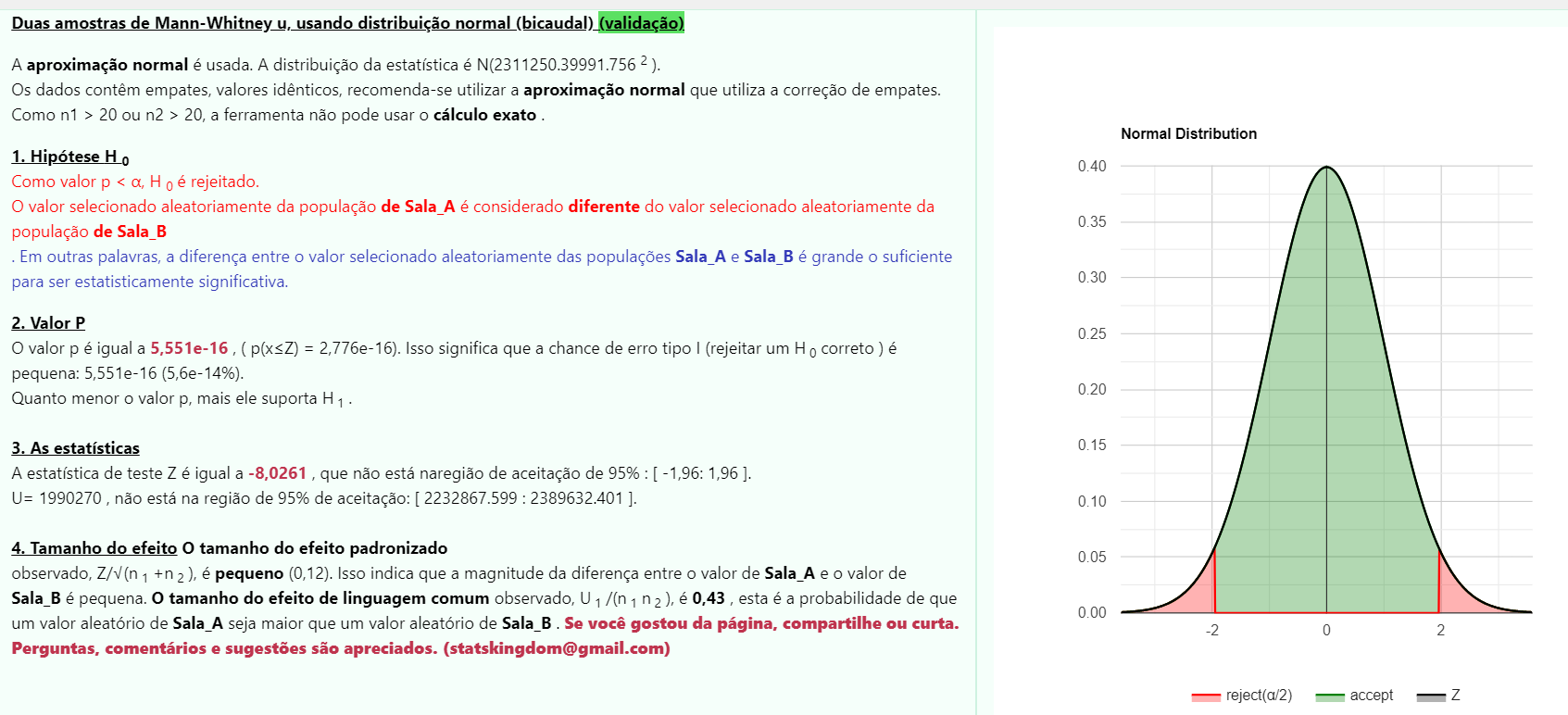
Após obter o resultado de normalidade dado através do modelo de shapiro wilk, podemos perceber que não há normalidade quando comparadas as salas A e B e suas zonas pertencentes que são 1 e 2.

Logo após isso fizemos uma pesquisa dentro do https://www.canalpesquise.com.br/tut-1tutorialestatistico-47/tut-45 para conseguirmos conduzir que tipo de análise deveríamos fazer com base nos nossos dados e na normalidade que no caso não há. Após preenchermos a pesquisa chegamos ao resultado de que o tipo de análise que iremos usar é o:



E com isso aplicamos o modelo de Mann-Whitney onde usamos os dados obtidos de ambos os grupos e chegamos a esse resultado:  
  


E ele nos dá juntamente com o modelo de Mann-Whitney a hipótese também:



Hipótese Nula (H0): Não há diferença significativa entre os valores selecionados aleatoriamente das populações das salas A e B.

Hipótese Alternativa (H1): Existe uma diferença significativa entre os valores selecionados aleatoriamente das populações das salas A e B.

Essa hipótese nula e alternativa são baseadas no resultado do teste de Mann-Whitney, onde a hipótese nula é rejeitada, indicando que a diferença entre os valores selecionados aleatoriamente das populações das salas A e B é estatisticamente significativa.

Com base na rejeição da hipótese nula no teste de Mann-Whitney, a hipótese alternativa (H1) é considerada verdadeira. Isso significa que há uma diferença significativa entre os valores selecionados aleatoriamente das populações das Salas A e B. Em outras palavras, as amostras das Salas A e B não tem a mesma distribuição, e essa diferença é estatisticamente significativa.

Com isso podemos concordar com a afirmação dos funcionários sobre a diferença de temperatura presente nas salas, e que há uma diferença significativa de temperatura entre as salas A e B, dependendo de sua zona e sala.