**Relatório de Análise de Dados de Aluguéis em São Paulo**

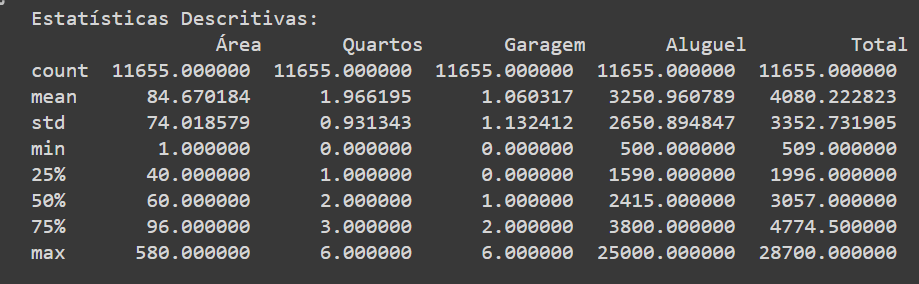
Este relatório apresenta os resultados da análise exploratória de dados de um conjunto de dados de imóveis para aluguel em São Paulo, disponível no link: <https://raw.githubusercontent.com/Leandrogbrito/Projeto-pratico-de-dados/main/base-alugueis-sp.csv>.

**1. Pré-processamento dos Dados**

* **Limpeza e Padronização:** O conjunto de dados passou por um processo de limpeza, incluindo:
  + Correção de endereços incompletos, separando informações como rua, bairro, cidade e estado.
  + Padronização da coluna 'Tipo', separando 'Studio e kitnet' em 'Studio' e 'Kitnet' com base no número de quartos.
  + Remoção de linhas com área igual a zero.
  + Remoção de colunas redundantes ou irrelevantes.
  + Renomeação de colunas para nomes mais descritivos em português.
  + Reorganização das colunas para uma ordem lógica.

**2. Análise Descritiva**

**Estatísticas Descritivas:**



**Observações:**

* A média da área dos imóveis é de 90,54 m², com um desvio padrão de 62,60 m².
* A maioria dos imóveis possui 1 ou 2 quartos.
* O aluguel médio é de R$ 3335,55, com uma variação significativa (desvio padrão de R$ 2532,17).

**Distribuição das Variáveis:**

* A análise dos histogramas mostra que a distribuição do aluguel é assimétrica à direita, com uma concentração maior de valores baixos e um número menor de valores altos.
* A distribuição do número de quartos também é assimétrica à direita, com maior concentração em imóveis com 1 ou 2 quartos.

**Correlação entre Variáveis:**

* A matriz de correlação indica uma correlação positiva moderada entre o número de quartos e o valor do aluguel.
* A área também apresenta uma correlação positiva moderada com o valor do aluguel.

**3. Modelagem de Regressão Linear**

* Um modelo de regressão linear foi criado para investigar a relação entre o número de quartos e o valor do aluguel.
* O modelo apresentou um **RMSE** (erro quadrático médio da raiz) de **1226,83** e um **R²** (coeficiente de determinação) de **0,52**.
* Esses resultados indicam que o modelo consegue explicar 52% da variabilidade do aluguel.

**Equação da Reta de Regressão:**

* Aluguel = 1266,64 + (1099,59 \* Quartos)

**Interpretação:**

* O modelo sugere que o valor do aluguel aumenta em média R$ 1099,59 para cada quarto adicional.
* O intercepto de R$ 1266,64 representa o valor estimado do aluguel para um imóvel com 0 quartos.

**Considerações:**

* A correlação entre o número de quartos e o aluguel é moderada, indicando que outros fatores também influenciam o valor do aluguel.
* O modelo de regressão linear deve ser utilizado com cautela, pois a análise exploratória não identificou outros fatores importantes que poderiam ser incluídos no modelo.

**4. Conclusões**

* A análise dos dados de aluguel em São Paulo revelou que o número de quartos é um fator significativo que influencia o valor do aluguel.
* O modelo de regressão linear apresentou uma capacidade moderada de prever o valor do aluguel com base no número de quartos.
* A análise exploratória destacou a necessidade de investigar outros fatores que podem influenciar o valor do aluguel, como localização, tipo de imóvel, área, etc.
* É recomendado realizar estudos adicionais para construir um modelo mais preciso que inclua outros fatores relevantes.

**5. Para ficar mais completo, seriam necessários os procedimentos a abaixo.**

* Realizar uma análise mais aprofundada da influência da localização e dos outros fatores na determinação do valor do aluguel.
* Considerar a aplicação de outros modelos de regressão, como a regressão múltipla, para incorporar mais variáveis ​​na predição do aluguel.
* Analisar a influência de variáveis categóricas, como o tipo de imóvel e o bairro, no valor do aluguel.
* Realizar uma análise de clustering para agrupar imóveis com características semelhantes e identificar tendências de preços.

Obs. Estou aberto a sugestões e criticas ao processo para uma melhor analise.

Quero aprender mais e cresce no conhecimento. E se útil para a sociedade com minhas analises.