



Projeto de Insights - House Rocket

Links rápidos

Login · Comunidade DS

<https://membro.comunitedatascience.com/37217-python-do-zero-ao-ds>

GitHub - renankalfa/python-0-ao-Data_Scientist: Curso de análise de dados para responder grandes e difíceis problemas.

You can't perform that action at this time. You signed in with another tab or window. You signed out in another tab or window. Reload to refresh your session. Reload to refresh your session.

https://github.com/renankalfa/python-0-ao-Data_Scientist

Checklist Geral

☒ Questão/Premissas de Negócio

☒ Planejamento da solução

☒ Coleta/Limpeza de Dados

☒ Exploração de Dados

☐ Criar um modelo com grandes e difíceis problemas.

☐ Documentar no GitHub

kalfa/python-0-ao-Data_Scientist

análise de dados para responder de negócio.

0 Issues 1 Star 0 Forks

1. Questão de Negócio

▼ Perguntas do CEO

P1. Quais são os imóveis que a House Rocket deveria comprar e por qual preço?

P2. Uma vez o imóvel comprado, qual o melhor momento para vendê-lo e por qual preço?

▼ Hipóteses

H1. Imóveis que possuem vista para água, são 20% mais caros na média.

H2. Imóveis com data de construção menor que 1955, são 50% mais baratos na média.

H3. Imóveis sem porão possuem área total (do lote), são 40% maiores do que os imóveis com porão.

H4. Imóveis com 3 banheiros tem um crescimento de MoM de 15%.

H5. O crescimento do preço dos imóveis YoY é de 10%.

2. Planejamento da Solução

2.1 Planejamento do Produto Final

▼ Perguntas do CEO

P1: Relatório com as sugestões de compra de imóvel por um valor recomendado.

P2: Relatório com as sugestões de venda de um imóvel por um valor recomendado.

▼ Hipóteses

H1-H5: Gráficos para a validação das hipóteses.

2.2 Planejamento da Solução

▼ Perguntas do CEO

▼ P1

▼ Plano 1

- Agrupar os dados por região (zipcode).
- Calcular a mediana por região.
- Irei recomendar imóveis que estão abaixo da mediana do preço de cada região e em boas condições.
- Criar um dataframe com os atributos: cod imóvel, região, preço, mediana do preço, condição e status.

▼ P2

▼ Plano 1

- Criar uma coluna com a sazonalidade.
Summer - março à agosto
Winter - setembro a fevereiro
- Calcular a mediana dentro de cada região e sazonalidade
- Condições de venda:
 - Preço da compra (price) for maior que a mediana da região + sazonalidade
 - Preço de venda será igual ao preço de compra + 10%
 - Preço da compra for menor que a mediana da região + sazonalidade
 - Preço de venda será igual ao preço de compra + 30%

Imóvel	Cod	Região	Temporada	Preço da Mediana	Preço da Compra	Preço Venda	Lucro
10330	302349	Verão		R\$ 800.000,00	R\$ 450.000,00	R\$ 450.000,00 + 30%	??
10330	302349	Inverno		R\$ 800.000,00	R\$ 450.000,00	R\$ 450.000,00 + 30%	??

▼ Hipóteses

▼ H1

- Comparar a média dos imóveis com vista pra água com os sem vista para água.
- Gráfico de pizza, barras (cor pela porcentagem).

▼ H2

- Comparar imóveis com yr_built menor que 1955 com os construídos a partir desta data.

- Gráfico de pizza, barras (cor pela porcentagem).

▼ H3

- Comparar imóveis com porão e sem porão pela área total do terreno. Como também o preço.
- Dashboard (): gráfico de pizza do tamanho do lote por ter porão ou não e gráfico de barras da média do preço de cada um.

▼ H4

- Pesquisar sobre "Month over Month".
- Filtrar por imóveis com 3 banheiros.
- Criar uma coluna que indica o ano/mês.
- Comparar o crescimento MoM.
- Gráfico de linha.

▼ H5

- Pesquisar sobre "Year over Year".
- comparar o crescimento dos anos.
- Gráfico de linha.

2.1 Planejamento das Ferramentas

- Python 3.9.6
- Python (Anaconda) 3.8.12
- Jupyter Notebook
- PyCharm

3. Exploração dos Dados

Ideias

- ▼ Hipóteses extras

Checklist

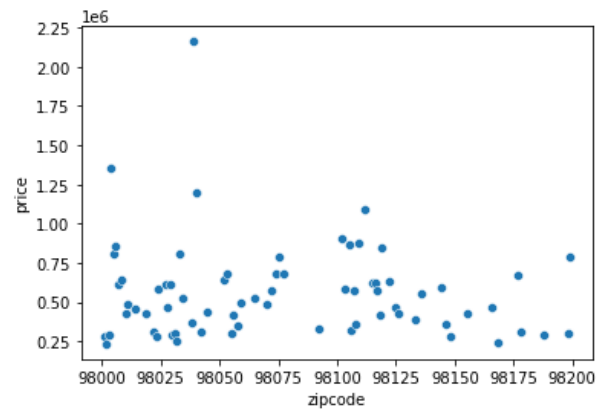
- ☒ P1
- ☒ P2
- ☒ H1-H3
- ☒ H4-H5

Análises Extras

- ▼ Zipcode outlier by price

```
d1 = data[['price', 'zipcode']].groupby('zipcode').mean().reset_index()
```

```
sns.scatterplot(x='zipcode', y='price', data=d1);
```



```
d1[d1['price'] >= 2160606.6]
```

	zipcode	price
24	98039	2160606.60

Ativar o Windows
Acesse Configurações para ativar o Windows.