

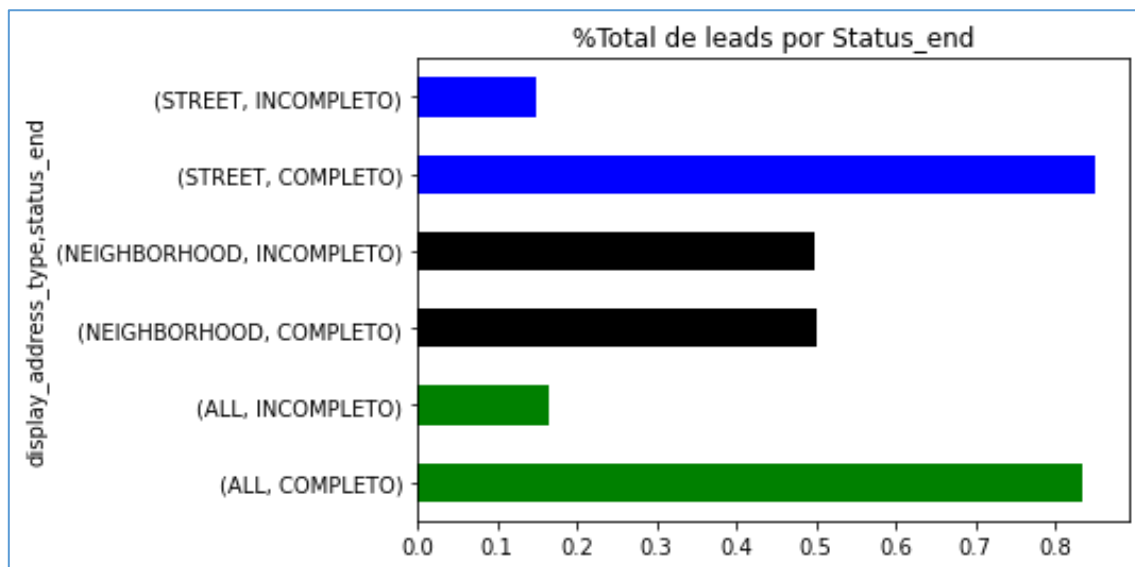
A análise foi realizada em python devido a facilidade de manipular os dados e tirar estatísticas da base a ser analisada. Para facilitar, utilizei o Google Colaboratory, um ambiente de notebooks Jupyter que não requer configuração e é executado na nuvem

Primeiro, verifiquei a quantidade de anuncios com o endereço completo, e cheguei que 75% dos anuncios estão completos.

```
status_end
COMPLETO    0.747723
INCOMPLETO  0.252277
```

Dentro desses tipo de endereços, verifiquei qual a porcentagem de leads distribuido de acordo com o tipo de endereço e, verifiquei que apenas neighborhood possuem a distribuição de leads de forma igual. Em Street e All, o porcentual de leads é bem maior em **endereço completo**.

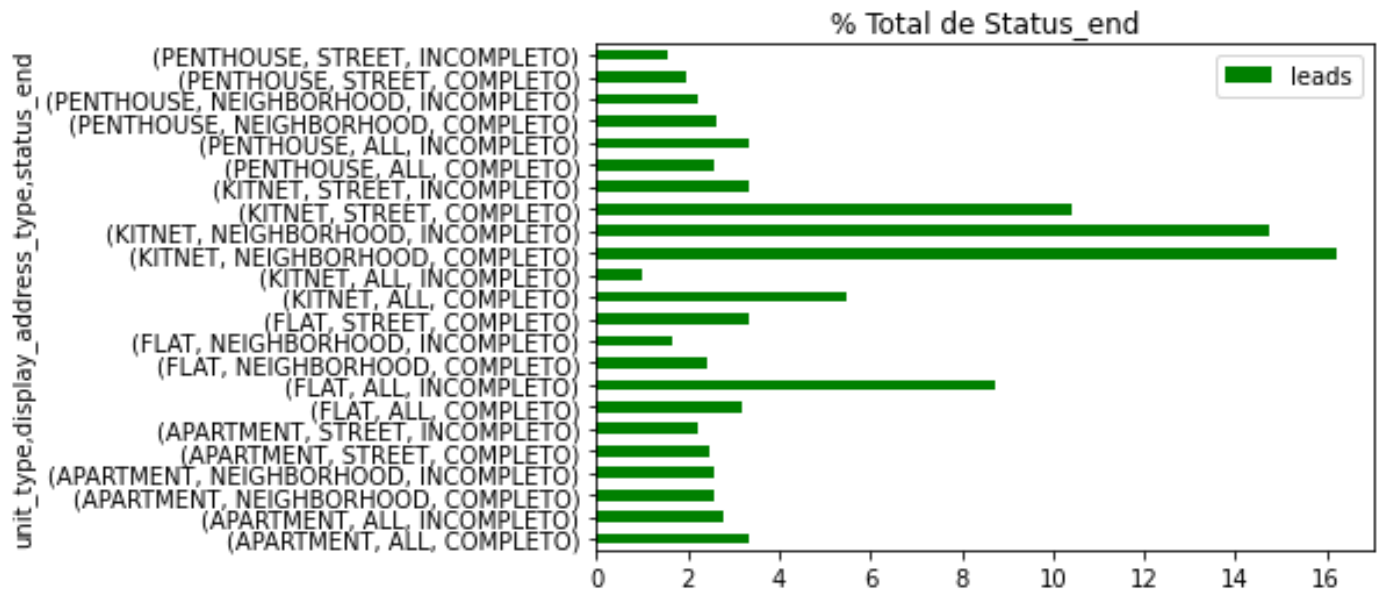
```
display_address_type status_end
ALL                  COMPLETO    0.834735
                   INCOMPLETO  0.165265
NEIGHBORHOOD        COMPLETO    0.501533
                   INCOMPLETO  0.498467
STREET              COMPLETO    0.850518
                   INCOMPLETO  0.149482
```



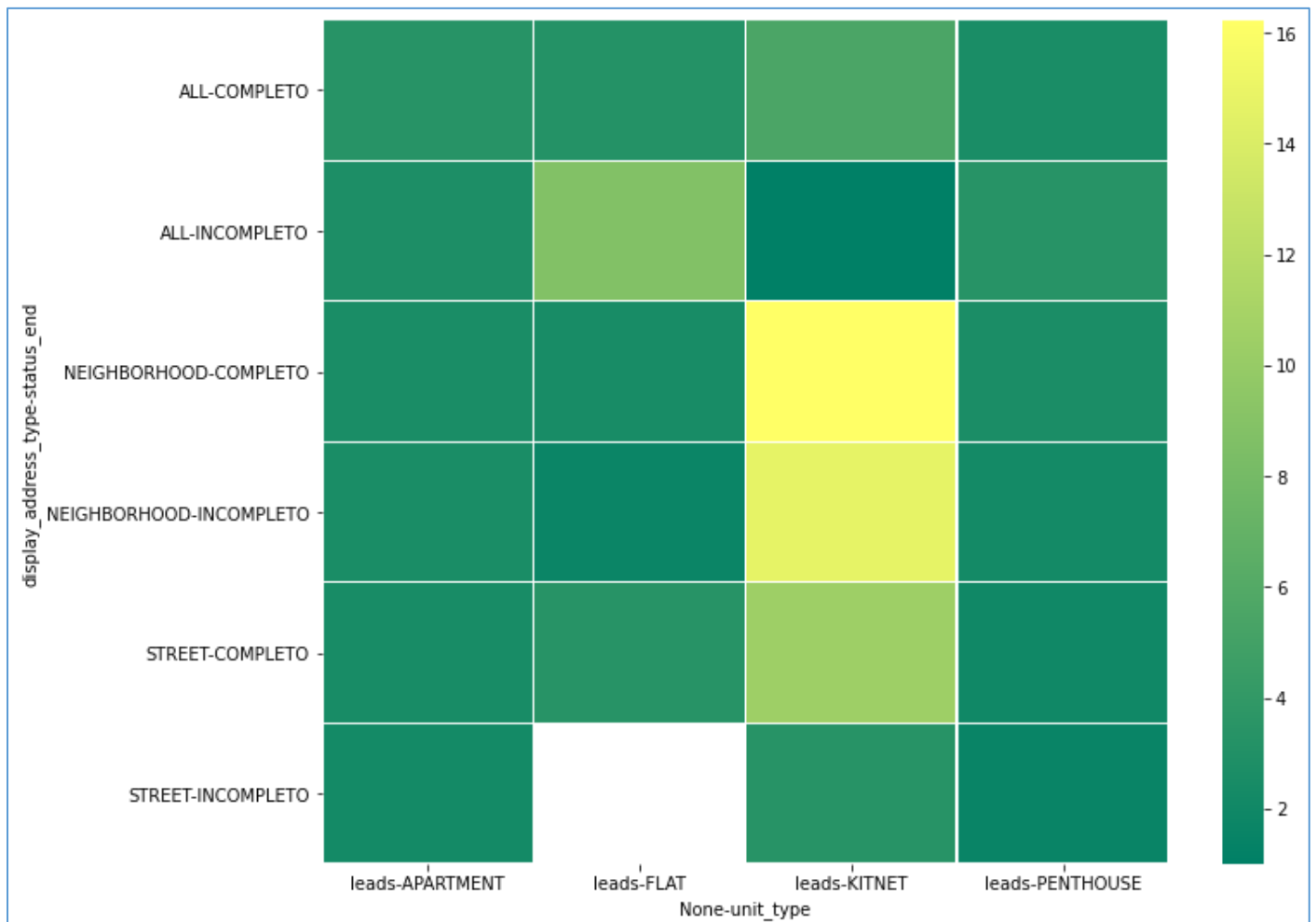
Se realizamos uma análise mais profunda ainda, levando em consideração a média de leads por tipo residência (unit_type) teremos a tabela abaixo. Vemos que somente um caso com uma média maior de visitas, em ALL > FLAT > Incompleto. Aonde verificamos uma diferença grande de média de leads por anuncio.

		leads			
	unit_type	APARTMENT	FLAT	KITNET	PENTHOUSE
display_address_type	status_end				
ALL	COMPLETO	3.352476	3.166667	5.473684	2.581081
	INCOMPLETO	2.770905	8.750000	1.000000	3.349057
NEIGHBORHOOD	COMPLETO	2.579165	2.444444	16.230769	2.640000
	INCOMPLETO	2.592157	1.666667	14.740741	2.245070
STREET	COMPLETO	2.461660	3.333333	10.421053	1.965293
	INCOMPLETO	2.206540	NaN	3.333333	1.542857

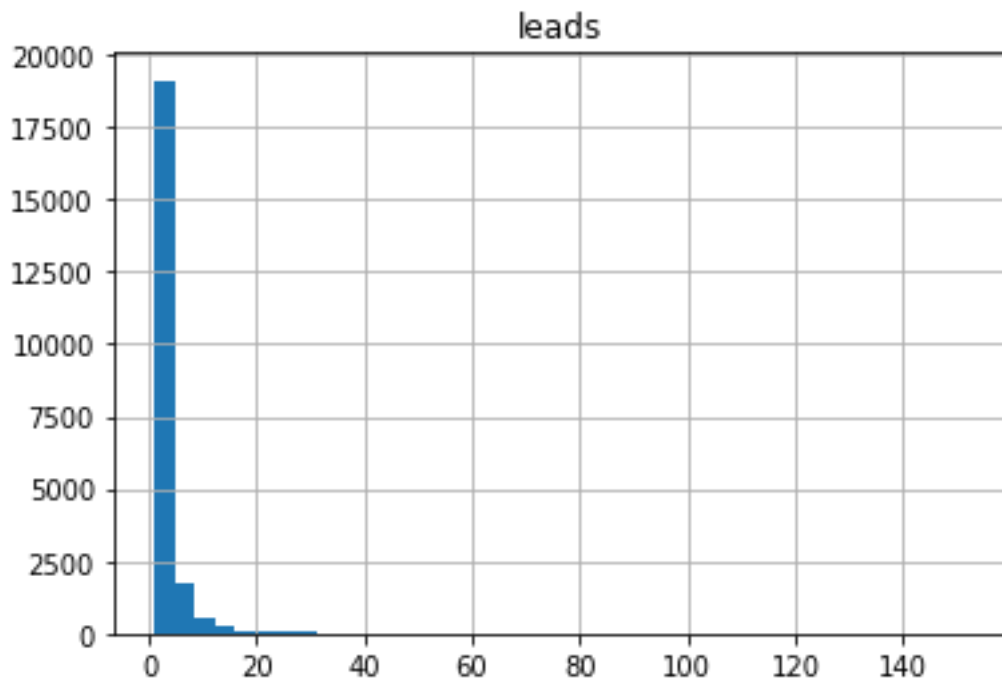
Plotando a tabela acima em um gráfico de barras, teremos a visualização abaixo.



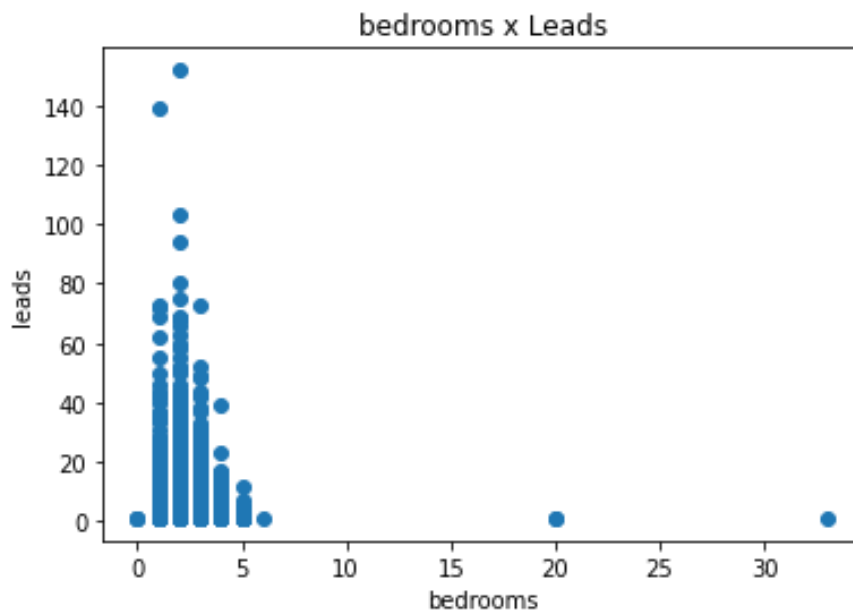
Entretanto, o gráfico possui muita informação e fica difícil comparar realmente os casos de acordo com o tipo de residência. Por isso, plotei o mapa de calor, que facilita a verificar de distorções entre os valores, e a comparação entre os diversos tipos.

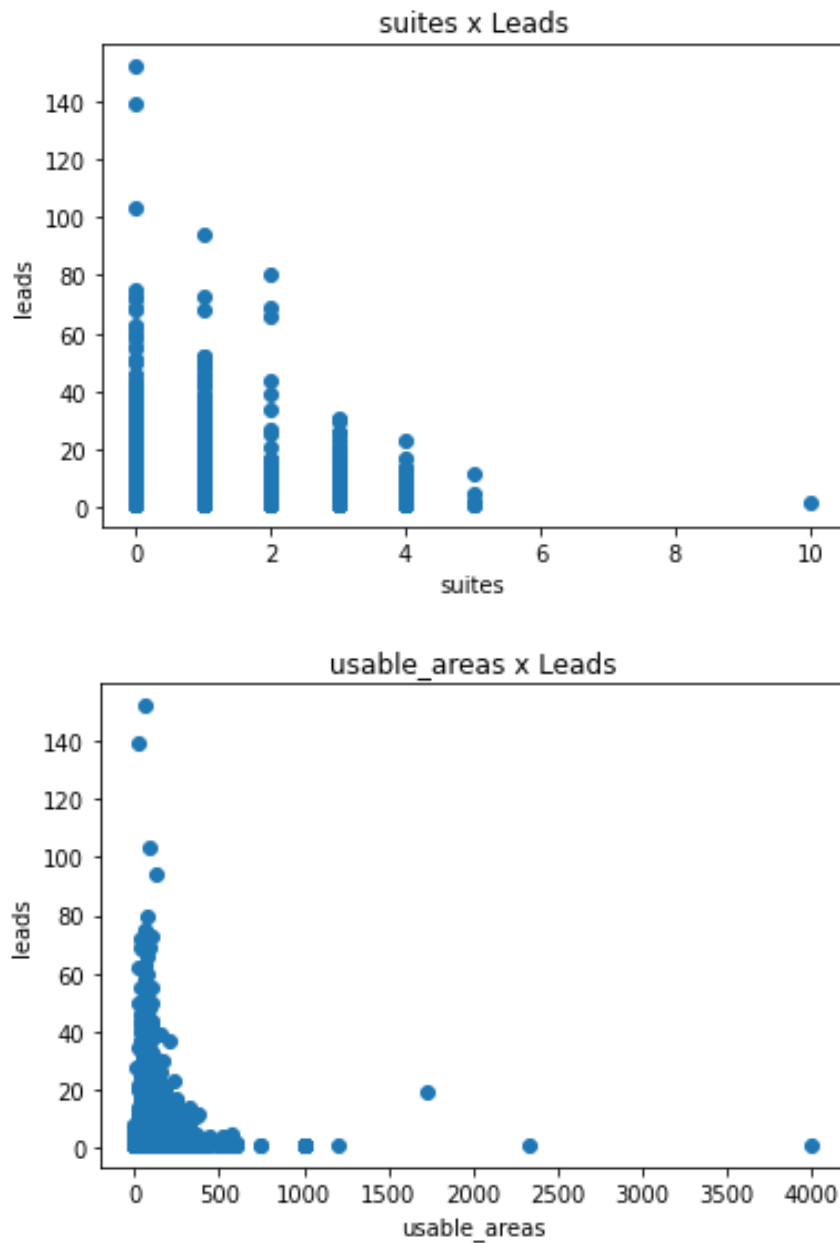


Realizando uma análise exploratória dos dados, realizei o histograma de leads, e verifiquei uma concentração de 1 a 5 leads.



Abaixo realizei a correlação dos leads com outras variáveis, para assim verificar se existe algum fator com maior influência e aonde está a maior concentração de leads.





Considerações

Portanto, verificamos que não é correta a premissa de que endereços incompletos atraem mais leads.

E verificamos que um anuncio com 1 ou 2 bedrooms, até 1 suíte e com usable área de até 100 m² é aonde estão os maiores números de leads por anuncio.

Finalizando, o anuncio com a maior média de leads (16) é a kitnet, com endereço completo e mostrando o bairro.

Melhorias

Falta um data wrangling para os bairros, existe muitos dados sem padrão (Paraíso e Paraíso). Falta saber se é anuncio de venda ou aluguel, somente pelo preço pode causar erros de classificação.