

Trabalho Prático 1

Desafio 1. Análise exploratória de um conjunto de dados de preços de imóveis

Introdução

A imobiliária [Properati](#) publica periodicamente informações sobre ofertas de imóveis para venda e aluguel. Deve aconselhar o agente imobiliário a desenvolver um modelo de regressão que permita prever o preço por metro quadrado de um imóvel. O objetivo final é que o modelo desenvolvido por eles seja utilizado como avaliador automático a ser aplicado nos próximos imóveis que forem comercializados pela empresa. Para isso, a empresa disponibiliza um conjunto de dados correspondente ao primeiro semestre de 2017.

O conjunto de dados é de tamanho pequeno a médio, mas tem duas complexidades às quais você precisa prestar atenção:

- Peso de dados faltantes em algumas variáveis relevantes.
- Será importante ter em conta o problema da influência espacial nos preços por metro quadrado. De fato, é provável que existam diferenças importantes nas diferentes geografias, bairros e áreas analisadas.

Objetivos:

- Executar uma limpeza do conjunto de dados fornecido. Em particular, você precisará elaborar estratégias para lidar com dados ausentes em determinadas variáveis.
- Realizar uma análise descritiva das principais variáveis.
- Criar novas colunas de determinados recursos que podem ter valor preditivo.

Datos

Baixe o conjunto de dados do Playground

Requisitos

Os materiais devem ser entregues em um notebook Jupyter que atenda aos requisitos do projeto. O notebook deve ser devidamente comentado. Além disso, os grupos precisarão criar um repositório para o projeto (anônimo) no Github. Para a apresentação, deve-se montar alguns slides não técnicos e gravar um vídeo que exponha os resultados obtidos em no máximo 10 minutos.

Material a entregar

Apresente um relatório estruturado (PPT ou Google Slides). Este deve consistir em uma introdução (enunciado do problema, a pergunta, a descrição do conjunto de dados, etc.), um desenvolvimento das análises realizadas (análise descritiva, análise de correlação preliminar, visualizações preliminares) e uma apresentação dos principais resultados e conclusões.

Fecha de entrega

- O material deverá ser entregue na **aula 18 do curso**.

Dataset

O conjunto de dados contém informações sobre todos os imóveis georreferenciados do banco de dados da empresa. A informação de cada propriedade que inclui é a seguinte:

Scheme : "Real_State_Properati.csv"

property_type : Type of property (House, Apartment, PH).

place_name : Name of the development.

country_name : Country in which the ad is published.

state_name: Fields referring to the location of the property.

lat : Latitude

lon : Longitude

price : Price published in the ad.

currency : Currency of the published price.

surface_total_in_m2 : Total area in m².

Como começar? Sugestões

Adicione qualquer informação adicional construída a partir dos dados originais (ou mesmo informações externas) que você considere relevantes e úteis para resolver os objetivos declarados.

Aproveite as ferramentas do Pandas: groupby, summation, pivot_tables e outros aplicativos e métodos de DataFrames tornam os cálculos e outras agregações de dados muito mais simples.

Ao apresentar os resultados, lembre-se de que é muito provável que o público não tenha um nível técnico, portanto, mantenha a linguagem em um nível acessível.

Em termos gerais, lembre-se das seguintes sugestões:

- Escreva um pseudocódigo antes de começar a escrever código. Muitas vezes é muito útil fornecer um esquema geral e uma lógica para a análise.
- Leia a documentação de qualquer tecnologia ou ferramentas de análise que eles usar. Às vezes não há tutoriais para tudo e os documentos e ajuda são essenciais para entender o funcionamento das ferramentas utilizadas.
- Documente todas as etapas, transformações, comandos e análises que eles realizam.

Recursos úteis

- [Github Help](#)
- [GeoPandas documentation](#)
- [Cheatsheet Pandas](#)
- [User guide Pandas](#)

Avaliação

/10	Discussão do dataset: análise exploratória dos dados brutos
/10	Discussão do método: Criando novas variáveis com valor preditivo

	Justificativa para a escolha de novas variáveis com valor preditivo
/10	Discussão do método: Detecção de campos a imputar Justificativa da seleção dos campos a imputar
/10	Discussão do método: Estratégia para obter um banco de dados de registros completos e com variáveis úteis para prever o valor de m2 dos imóveis. Justificativa da estratégia proposta para obter o conjunto de dados resultante
/10	Discussão dos resultados: análise exploratória do dataset resultado
/10	Profundidade de trabalho adequada
/10	Apresentação adequada das conclusões
/10	Formato do relatório (intro/corpo/conclusão, gráficos)
/10	Apresentação do relatório (estrutura do texto, narrativa, figuras)
/10	Apresentação do vídeo (tempo adequado, clareza na apresentação, informações relevantes para o público definido)
/100	Nota final