

Case Técnico

Data Science

Labs

palpe.co

Descrição do problema:

Uma empresa japonesa que atua no mercado imobiliário recebeu uma base de dados do MLIT (Ministry of Land, Infrastructure, Transport and Tourism of Japan), e a partir disso busca criar um modelo que faça previsão do valor de mercado dos imóveis na região. Para isso, a empresa contratou você para desenvolver esse projeto.

Descrição dos dados:

O dataset é referente a transação de imóveis em Tokyo, no período de 2005 à 2019, contendo diversas informações, conforme a tabela abaixo:

Coluna	Tipo	Descrição	Exemplo
No	INTEGER	ID do imóvel	12345
Type	TEXT	Tipo de imóvel	Residential Land (Land Only)
Region	TEXT	Características das áreas circundantes	Commercial Area
MunicipalityCode	INTEGER	Código da cidade no Japão	13101
Prefecture	TEXT	Nome da prefeitura no Japão	Tokyo
Municipality	TEXT	Cidade	Adachi Ward
DistrictName	TEXT	Distrito	Zushimachi
NearestStation	TEXT	Nome da estação mais próxima	Kobuchi
TimeToNearestStation	TEXT	Descrição temporal até a	10-15min

		estação mais próxima	
MinTimeToNearestStation	INTEGER	Tempo mínimo até a estação mais próxima (minutos)	10
MaxTimeToNearestStation	INTEGER	Tempo máximo até a estação mais próxima (minutos)	15
FloorPlan	TEXT	Planta do imóvel	2LDK
Area	INTEGER	Área total do imóvel (m ²)	220
AreaIsGreaterFlag	INTEGER	Indica se a área é menor ou igual à 2000 m ² (0) ou maior (1)	0
UnitPrice	REAL	Preço por unidade de área (Yen por m ²)	95000.0
PricePerTsubo	REAL	Preço por Tsubo (~3.3 m ²) (Yen por Tsubo)	320000.0
LandShape	TEXT	Forma do terreno	Semi-rectangular Shaped
Frontage	REAL	Área da Fachada (m ²)	11.5
FrontageIsGreaterFlag	BOOL	Indica se a área da fachada é ≤ 50m ² ou > 50m ²	False
TotalFloorArea	INTEGER	Área total do piso (m ²)	10
TotalFloorAreaIsGreater	INTEGER	Indica se a área total do piso é menor ou igual à	1

		2000 m ² (0) ou maior (1)	
BuildingYear	INTEGER	Ano de construção do edifício	1979
PrewarBuilding	INTEGER	Indica construções feitas antes (0) de 1945 ou depois (1)	0
Structure	TEXT	Tipo de estrutura do edifício*	LS
Use	TEXT	Uso atual do imóvel	House
Purpose	TEXT	Propósito de uso futuro	Shop
Direction	TEXT	Direção da fachada	Northwest
Classification	TEXT	Classificação da via em frente	Prefectural Road
Breadth	REAL	Largura da via em frente (m)	3.7
CityPlanning	TEXT	Uso dos distritos designados pela Lei de Planejamento Urbano	Urbanization Control Area
CoverageRatio	INTEGER	Percentual máximo de cobertura do edifício (%)	80
FloorAreaRatio	INTEGER	Percentual máximo da área do piso (%)	300
Period	TEXT	Tempo da transação	2nd quarter 2016
Year	INTEGER	Ano da transação	2016

Quarter	INTEGER	Trimestre do ano da transação	2
Renovation	TEXT	Reforma?	Not yet
Remarks	TEXT	Observações	Dealings including private road
TradePrice	INTEGER	Preço da transação do imóvel (Yen)	120000000

*SRC= Steel frame reinforced concrete, RC= Reinforced concrete, S = Steel frame, LS = Light steel structure, B = Concrete block, W = Wooden

Dados de treino e teste:

- Serão fornecidos dois datasets para os candidatos:
 - Dados de treino: contém todas as colunas (incluindo a variável alvo – TradePrice)
 - Dados de teste: contém as colunas, mas sem a variável alvo. **Será exigido o envio desse dataset com as previsões do modelo**, conforme a explicação na seção “O que será exigido?”

O que será exigido ?

- Subir o projeto no GitHub/GitLab/BitBucket;
- Criar base de dados SQLite/Postgres/MySQL/... com os dados disponibilizados;
- Carregar os dados dessa base para um Jupyter Notebook (é nesse notebook que será desenvolvido o projeto);
- Preparação e limpeza de dados;

- Processos que você julgue necessário;
- Análise Exploratória;
 - Apresentar pelo menos dois insights sobre os dados (De preferência de forma visual com um gráfico);
- Criar um modelo de Predição para a variável *TradePrice* (O modelo usado fica a seu critério, basta justificá-lo);
- Medir a performance do modelo com as métricas RMSLE e MAPE;
- Após a finalização do modelo, **enviar o dataset de teste (em formato .csv) com as previsões do modelo para a variável *TradePrice*.**
- Gravar um vídeo de até 10 minutos, mostrando os resultados e explicando brevemente cada passo realizado na solução;
- Criar alguma visualização para avaliar os resultados das previsões do modelo.

Portanto, o que deve ser enviado no final?

- Link do repositório, contendo o notebook com o desenvolvimento do projeto;
- Vídeo com as explicações;
- CSV com as previsões feitas nos dados de teste;