

Trabalho de grupo IMD

Licenciatura em Ciência de Dados (2º ano)

03 de Março de 2024

Diana Mendes (diana.mendes@iscte-iul.pt)

Conceição Figueiredo (conceicao.figueiredo@iscte-iul.pt)

Ricardo Correia (ricardo_correia@iscte-iul.pt)

Deadline entrega: 14 de abril de 2024

Tema: Previsão dos preços em reservas em hotéis/alojamentos - desenvolver um modelo que permita estimar o preço a que os quartos são colocados no Airbnb à venda (<http://insideairbnb.com/get-the-data/>).

- Os docentes atribuem a cada grupo uma cidade (ver em **Anexo**)
- Os dados podem ser importados do site **Inside Airbnb**, <http://insideairbnb.com/get-the-data/>
- Apenas fazer o download do ficheiro [listings.csv](#) da cidade do seu grupo.
- Cada ficheiro tem milhares de linhas e dezenas de colunas (variáveis)
- Tirar tempo para procurar o máximo de informação de contexto que permita a **compreensão** dos **dados** e do **domínio** de onde provêm.
- Neste trabalho recomenda-se usar o software R (script em RStudio ou Notebook em Jupyter Notebook, IRkernel).

Pontos principais:

O desenvolvimento do projeto e do relatório final deve seguir a metodologia CRISP-DM.

1. Com base nos dados e a sua pesquisa sobre os dados (*Business Understanding*), definir o(s) *problema(s)/questões* que gostariam de *resolver/responder* e justificar a escolha das variáveis.
2. Fazer a limpeza dos dados.
3. Analisar gráficos e estatísticas descritivas das variáveis.
4. Analisar correlação e causalidade entre as variáveis.
5. Pré-processamento dos dados e manipulação de *features* (agrupar, juntar, eliminar, transformar as variáveis).
6. Usar algoritmos de aprendizagem supervisionada (regressão linear, regressão polinomial, interação de variáveis, regressão não-linear) sobre o seu conjunto de dados.
7. Dividir o dataset em conjuntos de treino e de teste.
8. Validar o modelo escolhido e fazer a previsão da variável dependente/alvo (sobre o conjunto de teste).
9. Avaliar a performance da previsão feita (sobre o conjunto de teste).
10. Interpretação/explicação dos resultados obtidos a partir dos dados (prós e contras).
11. **Relatório:** podem usar *dashboards*, editores de texto, ou outras ferramentas.
12. O relatório final deve ser enviado por e-mail em **formato pdf**. Devem ainda submeter um ficheiro zippado com o dataset final e o ficheiro com o código utilizado.
13. O **código** tem de poder ser executado AS-IS (diretamente) e deverá estar devidamente comentado.
14. Deverá incluir todas as experiências feitas, incluindo a análise e preparação de dados, modelação, e avaliação do modelo.
15. A interpretação dos resultados deve ser efetuada de forma crítica e não apenas factual.
16. O relatório tem um **limite de 25 páginas**.

Anexo: Grupos e Cidades

Grupo	Nome	Nº de aluno	Turma	Email
1 Albany	Luana Pinto Tomás Mendes Nuno Matos Dinis Cardoso Catarina Raminhos	111145 110612 111102 105663 111095	CDB1	lbsmp@iscte-iul.pt tmms3@iscte-iul.pt nmbcm@iscte-iul.pt dacco1@iscte-iul.pt cclcr@iscte-iul.pt
2 Austin	Maria Ferreira Catarina Lameira Diana Rodrigues Margarida Salgueiro	110986 111390 111663 111566	CDB2	mmsfa1@iscte-iul.pt cclao@iscte-iul.pt dnrss@iscte-iul.pt mibso@iscte-iul.pt
3 Valencia	Carolina Brunheta Luís Ferreira Leonardo Dias Duarte Alcobia Camila Sousa	110888 110854 110752 111309 111017	CDB1	cvpsb@iscte-iul.pt lphfa@iscte-iul.pt lvbpd@iscte-iul.pt dmetaa3@iscte-iul.pt ctsas@iscte-iul.pt
4 Bordeaux	Gonçalo Girão Eduardo Parracho Tomé Marques Gabriel Matos Diogo Aqueu	111515 111423 110966 110907 110705	CDB2	gagoa@iscte-iul.pt efspo@iscte-iul.pt tmms2@iscte-iul.pt gomsa1@iscte-iul.pt drcau@iscte-iul.pt
			CDB1	
5 Victoria	Carolina Santos Luana Guilherme Madalena Sebastião Nuno Gonçalves	110815 111153 110844 106365	CDB1	canrs1@iscte-iul.pt lagg@iscte-iul.pt macso1@iscte-iul.pt nmags@iscte-iul.pt
6 Copenhagen	Afonso Lourenço Afonso Santos Diogo Santos Guilherme Simões Afonso Pereira	111487 111431 111386 111604 111134	CDB2	axsnl@iscte-iul.pt aassa3@iscte-iul.pt dlss2@iscte-iul.pt gnssa@iscte-iul.pt atpao@iscte-iul.pt
			CD-PL-B1	
7 Denver	Rodrigo Casimiro João Copeto João Magarça Sara Marques	111177 110867 111640 111332	CDB1 CDB1 CDB2 CDB1	jmsma4@iscte-iul.pt
8 Geneva	Alexandra Dovgan Lurdes Oliveira Marta Cardita Marta Pedro	106845 110613 111292 111382	CDB1	aadna@iscte-iul.pt lmpo@iscte-iul.pt msmca2@iscte-iul.pt mgpoo@iscte-iul.pt

	Duarte Salvador	111181	CDB1	dgssr@iscte-iul.pt
9 Lyon	Margarida Lopes	110372	CD-PL-B1	mjdmc@iscte-iul.pt
	Matilde Brito	110794	CDB1	msbos2@iscte-iul.pt
	Mariana Estêvão	111364	CD-PL-B1	mpfeo2@iscte-iul.pt
	Rodrigo Ramos	11360	CD-PL-B1	rssrs3@iscte-iul.pt
10 Milan	Mafalda Golaio	111067	CD-PL-B1	madgo1@iscte-iul.pt
	Pedro Paiva	110577		Pmgpa@iscte-iul.pt
	Tiago Pacheco	111120		tmpos1@iscte-iul.pt
	Joana Felizardo	106842		jsmcf1@iscte-iul.pt
	Marta Facaia	111304		mpsfa3@iscte-iul.pt
11 Montreal	Inês Anahory	123817	CDB2	insay@iscte-iul.pt
	Leonor Laborinho	111287		latfl@iscte-iul.pt
	Filipe Rego	111533		fwcro@iscte-iul.pt
	João Madeira	105188		jgbma@iscte-iul.pt
12 Zurich	João Pina	105855	PL	jbmpa1@iscte-iul.pt
	Afonso Couto	112136		apbfc@iscte-iul.pt
	Enzo Matavele	105747		emlme@iscte-iul.pt
	António Teotónio	111283		
	João Batista	111611		
13 Oakland	Leonor Bértolo	111141	CDB2	lagbo@iscte-iul.pt
	Bernardo Dinis	111401	CDB1	bmbvd@iscte-iul.pt
	Constança Brás	110682		cmpav@iscte-iul.pt
	Ana Rita Rodrigues	110823		arcrs2@iscte-iul.pt
	Mafalda Simões	110607		mvcss@iscte-iul.pt
14 Ottawa	João Dias	110305	CDB1	jprds@iscte-iul.pt
	Rafael Cerqueira	110860		rrcao2@iscte-iul.pt
	Samuel Ricardo	110884		svrom@iscte-iul.pt
	David Franco	110733		damfo1@iscte-iul.pt
	Felipe Pereira	110861		flpas@iscte-iul.pt
15 Prague	Rita Guerreiro	112018	CDB2	rpcgo@iscte-iul.pt
	Mariana Pimenta	111664		mfppa1@iscte-iul.pt
	Miguel Celestino	111590		maaco2@iscte-iul.pt
	Ana Almeida	111658		assaa6@iscte-iul.pt
	Filipe Duarte	104646		fdbde@iscte-iul.pt
16 Riga	Tomás Ferreira	104590	CDB1PL	tsfas2@iscte-iul.pt
	Ana Rodrigues	110830		asprs1@iscte-iul.pt
	Ariana Meiavia	106579		asmaz@iscte-iul.pt
	Tiago Reis	111424		
17 Rotterdam	Diogo Cardoso	105396	CDB1PL	nmnne@iscte-iul.pt tacoo1@iscte-iul.pt
	Nuno Norte	111505		
	Tomás Carvalho	111140		
	Xavier Simões	104605		
18	Francisco Inácio	111183	CDB2	flloo@iscte-iul.pt

San Diego	Ana Marta de Oliveira Mariana Cardoso Matilde Lemos	111066 111457 104935		amloa@iscte-iul.pt mlcas2@iscte-iul.pt
19 Sicily	Filipe Hristovsky Tomás Monge Francisco Coelho	93949 98574 111250	CDB1 CDB1 CDB1	fchyo@iscte-iul.pt fmgac@iscte-iul.pt
20 Tasmania	Vasco Solá Guerreiro Carlota Brejo Diogo Cancela António Coutinho	105301 98573 99369 98584		
21				