	Curso:	MBA em Tecnologias e Aplicações de Business Intelligence		
	Disciplina:	Técnicas e Programação de Analytics		
	Período	2019	Professor	LUÍS PORCIÚNCULA
	Trabalho Final da Disciplina			

Composição

Trabalho em equipe com o máximo de 04 componentes.

Propósito do Trabalho

Avaliar a capacidade dos alunos de preparar e interpretar dados para a obtenção de insights com dados que circundam um problema de Análise de Dados e Interpretação Analítica, com o apoio de um algoritmo proposto de análise preditiva para tomada de decisão.

O Problema a ser estudado

Uma seguradora de veículos quer melhorar o posicionamento de seus veículos de apoio (reboques e motos) nos períodos chuvosos para ter agilidade e fidelizar mais clientes em Salvador e contratou vocês para criar um algoritmo preditivo que a auxilie a tomar a melhor decisão sobre onde posicionar estes veículos nas regiões da cidade. E para isto ela deseja conhecer qual a característica do trânsito de cada bairro quando chove e quando não chove.

Após uma reunião interna na seguradora, decidiu-se que a estação do ano deve ser considerada para a análise, pois é de conhecimento que há uma variação da quantidade de chuva que cai em cada estação do ano. O gerente de marketing da empresa disponibilizou arquivos com informações sobre a quantidade de chuva que caiu na cidade e foi registrada pelos órgãos CEMADEN e INMET. Os arquivos são: [cemaden_ssa_01_2014_03_2019.zip](#) e [inmet_08_1963_03_2019.zip](#).

A empresa trabalha com sua própria regionalização da cidade, dividindo-a em 6 regiões: Aeroporto, Subúrbio, Cidade Baixa, Federação, Centro e Brotas e forneceu um arquivo chamado [seg_log_bairro_regiao_ssa.txt](#) que contém o nome dos bairros que compõem cada região.

A seu pedido, a seguradora também disponibilizou um arquivo com a divisão oficial dos bairros de Salvador que está disponível no arquivo [ra_bairros_pms.csv](#) para que você possa verificar se todos os bairros da cidade foram contemplados da regionalização feita pela seguradora e também arquivos com informações históricas sobre o trânsito na cidade, disponíveis nos seguintes arquivos:


[Seg_momentum.csv](#) – este arquivo contém a segmentação dos horários do trânsito na cidade.

O arquivo [dicionário_dados.pdf](#) contém informações sobre os campos dos arquivos disponibilizados.

Seu Objetivo

Produzir uma **análise preditiva básica** para o problema da seguradora de veículos através da geração de uma **árvore de decisão com a interpretação básica da mesma**, considerando o seguinte:

- **Analisar os dados** contidos nos arquivos que foram disponibilizados (CSV, TXT) como Datasets.

	Curso:	MBA em Tecnologias e Aplicações de Business Intelligence		
	Disciplina:	Técnicas e Programação de Analytics		
	Período	2019	Professor	LUÍS PORCIÚNCULA
	Trabalho Final da Disciplina			

- Avaliar estes arquivos e realizar a preparação de um ou mais datasets que possa(m) ser utilizado(s) para a construção do(s) Data Frames que serão utilizados com **fonte de dados** do script em R ou em Python que foi disponibilizado como template.
- A preparação da fonte de dados poderá ser feita utilizando uma das linguagens (R ou Python) ou utilizando outras ferramentas, como por exemplo o Excel, Ferramentas ETL, outras.
- Produzir e entregar a análise de no mínimo dois bairros.
- A análise são seus insights sobre o modelo criado por você com base nos dados processados pelo modelo que você criou;
- A análise deve ser entregue num documento PDF contendo sua interpretação levando em conta o que você considerou como importante para a sua avaliação. Deve estar disponível no documento o gráfico da árvore.
- Também deverá ser entregue o código do script com o(s) dataset(s) que ele processa.
- Prazo, até 29/07/2019. Enviar para luis.porciuncula@gmail.com

Dicas

- Os modelos preditivos seguem uma análise com base na função genérica: $y \sim X$, onde lemos y em função de X. Isto significa que o resultado do problema tem por característica avaliar o comportamento do trânsito em função da chuva, levando em consideração fatores que podem ou não enriquecer sua análise, como horário, estação do ano, velocidade das vias, etc.
- O foco é uma análise básica. Todo modelo analítico começa com uma análise básica, sem muitos fatores, por isso não complique, pois o modelo ideal começa da análise básica.
- Busque ter insights com as colunas existentes nos arquivos. No template Python tem um tipo insight provocado pela inclusão de uma nova coluna no Data Frame.
- Caso use Python, esteja atento(a) à necessidade de discretizar valores String para o algoritmo, pois o template está utilizando a Scikitlearn que não trabalha com dados não numéricos.