# Centro Universitário Instituto de Educação Superior de Brasília

Departamento de Pós graduação Especialização em Big Data, BI e Analytics Aplicados aos Negócios

## Projeto Integrador I

Aluna: Mariana Borges de Sampaio Professor: William de Almeida Silva

Disciplina: Fundamentos de Big Data e Conhecimentos de Dados

# Centro Universitário Instituto de Educação Superior de Brasília

Departamento de Pós graduação Especialização em Big Data, BI e Analytics Aplicados aos Negócios

## Projeto Integrador I

Primeiro relatório do Projeto Integrador 1 da disciplina de Fundamentos de Big Data e Conhecimentos de dados.

Aluna: Mariana Borges de Sampaio Professor: William de Almeida Silva

### Conteúdo

1	Introdução	1
2	Dicionário de dados e Análise de dados2.1 Dicionário de dados	
3	Considerações finais	3
$\mathbf{B}^{\mathbf{i}}$	ibliografia	4
$\mathbf{A}$	nexo	5

### 1 Introdução

Ao longo dos anos as empresas de todos os países e de todos os continentes foram se desenvolvendo, cada um em sua área específica, sejam empresas de corretoras de seguro, empreiteiras, empresas de tecnologia, empresas de saúde, entre os mais diversos tipos e variados setores em que uma empresa pode se aplicar. Apesar dessas empresas terem conceitos e fundamentos diferentes visto que elas são pertencentes a mundos diferentes, tem-se detém um conjunto de dados, esse dado muitas vezes pode não ser estruturado, sendo ele um dado bruto, um dado que só teria sentido se fosse tratado a fim de se tornar uma informação.

Sendo assim, cada empresa existente até os dias de hoje inevitavelmente possui dado, logo possui informação. Esse dado pode ser ele armazenado em um banco de dados estruturado ou pode ser um big data, sendo este um dado não estruturado. O big data tem por definição ser um conjunto de dados que possuem uma maior variedade, maior volume e estes dados têm uma maior velocidade. Sendo assim, o seu armazenamento é diferente do que um dado estruturado. Para isso conforme foram-se passando os anos foi sendo necessário nos meios do trabalho ter profissionais que são voltados para essa área.

### 2 Dicionário de dados e Análise de dados

#### 2.1 Dicionário de dados

A fonte de dados em questão foi extraída do Kaggle, conforme o que foi passado na dsecrição da atividade ativa.

Ao analisar o documento pode-se perceber a existência das seguintes colunas:

- id:
- roomid/id;
- noteddate;
- temp;
- out/in.

A seguir, tem-se a coluna com seu respectivo significado.

- id corresponde à um valor alfanumérico que indica uma identificação única, sendo assim cada linha é um momento em que foi registrada a temperatura do edifício;
- roomid/id corresponde ao nome da sala do edifício em que foi medida a temperatura. Para a identificação das medidas, tem-se o id que é um alfanumérico, visto que esses casos foram todos registrados em momentos aleatórios.
- noteddate corresponde a data de anotação que foi identificada a temperatura no edifício, sendo está registrada em intervalos aleatórios;
- temp corresponde a temperatura que foi atingida no edifício;
- out/in indica se a classificação de cada se a temperatura medida foi no exterior ou no interior do edifício.

#### 2.2 Análise de dados

Para realizar a análise de dados, foi necessário entender um pouco mais sobre o contexto do arquivo que foi extraído do kaggle. Os dados apresentados correspondem a dados oriundos de um sistema de monitoramento de temperatura da sala do edifício empresarial (admin), tanto no exterior como no interior do edifício.

3 Considerações finais

## Bibliografia

AGUIRRE, L. A. Introdução à Identificação de Sistemas, Técnicas Lineares e Não lineares Aplicadas a Sistemas Reais. Belo Horizonte, Brasil, EDUFMG. 2004.

### Anexo