**Desafio de Grupo nº 3-4**

**(até 3 de Dezembro)**

Vamos começar um novo projeto que começa com o seguinte material

1. Um ficheiro CSV: **20081026094426.csv** extraído de um Dataset **Geolife.**
2. Um User Guide que descreve os ficheiros desse Dataset.

**Tarefa 1.** Proponho-vos, portanto, a análise deste ficheiro usando ferramentas de extração de dados feitas por vocês em linguagens como **C, Phyton** ou outra ou utilizando ferramentas de análise de ficheiros csv tais como **CSV Viewer** ou outro.

Estes dados serão aplicados mais tarde noutras tarefas.

**Tarefa 2**.

Proponho que tentem ver:

**A)** **discrepâncias entre os dados**, **missing values** e formas de os corrigir. Procurem ferramentas de gratuitas de **“Limpeza de Dados”.**

**PS:** Normalmente, como os dados foram obtidos com smartphone poderão ter falhas devido a muitos fatores. Tentem adivinhá-los

**B)** Depois de “limpos” passe-os para uma folha de cálculo e calculem:

1. **Tempo decorrido de em segundos** de ponto para ponto da trajetória.
2. **Distância entre 2 pontos**, tendo em conta a esfericidade da terra. Usem a **fórmula de Haversine**.
3. Calculem a **velocidade de deslocação** entre pontos em m/s e km/h.
4. Infiram o **meio de transporte utilizado** em cada um deles.
5. Calculem a **distância total percorrida** e o **tempo total gasto**.
6. **Tirem conclusões** sobre a trajetória e os diferentes pontos.

**Tarefa 3.**

Tendo um novo ficheiro que vos enviei: 15-07-2017 ….. csv. Calculem tudo o que têm de fazer e calcular de acordo com a **Tarefa 2**.

Por outras palavras, nem sempre têm o mesmo ficheiro para a analisar, logo parametrizem toda a **tarefa 2** de modo a aceitar qualquer tipo de ficheiro.

**Tarefa 4.**

Partindo do princípio que podemos ter de analisar todo o tipo de dados, desde que possuam campos obrigatórios tais como:

1. Latitude
2. Longitude
3. Data
4. Time

Podem estar em colunas diferentes, ter linhas antes que são só de informação, pode ter ou não cabeçalho.

**Proponho:** **Caso aceitem**, um novo desafio em que vão aproveitar o que já têm feito que se resume a:

* Um menu em que podemos escolher:

1. O nome do ficheiro que queremos analisar
2. As colunas em que se encontram os dados referidos (Latitude, Longitude, Date, Time)
3. Quantas linhas são para desprezar
4. Em que linha está o cabeçalho (se não houver criá-lo ou apresenta-lo)
5. Fazer os estudos definidos anteriormente sobre estes dados.

O programa deve ser corrido (ou do ambiente de trabalho ou IDE) de modo a que a pessoa que quer analisar os dados possa proceder em conformidade.