

Etapa 2 – Preparação dos dados:

- Selecionar os Dados
- Limpar/ Uniformizar os dados
- Derivar dados
- Integrar os Dados
- Formatar os Dados

Todas as etapas foram descritas dentro código

Entrega2_Preparar_Dados_2204.ipynb nesse documento trazemos um breve resumo:

Seleção dos dados a serem utilizados:

Ride: RideID, Schedule, Create, Updated;

Ride Address: Number, Lat, Lng, RideAddressTypeID, RideID;

Ride Estimative: RideID, ProductID e Price;

Base “**Product**” desconsiderada.

Limpeza e uniformização dos dados:

Filtragem dos serviços Uber e 99;

Exclusão de valores nulos;

Conversão de datas.

Derivação de dados:

Derivação das datas em ano e mês;

Derivação de dias da semana;

Derivação de datas padronizadas

Derivação de origem e destino.

Formatação e integração:

Nomes de colunas renomeadas (lat, lng, number, rideaddresstypeid);

Merge de dados com base em rideid desconsiderando linhas no qual não possuam o rideid no DataFrame;

Ordenação por meio de rideid e price.

Concluindo:

Essa etapa trouxe muitos desafios e problemas a serem contornados (como mencionado no arquivo ipynb), porém esses desafios fizeram com que ganhássemos mais “intimidade” com a base de dados no qual iremos trabalhar daqui para frente.

Após essa etapa possuímos uma base bem estruturada que vai permitir avançarmos nas outras frentes do projeto não só em Ciência de Dados e Big Data como em IA e Machine Learning.