**Análise da Base de Dados de Celulares Subtraídos - 2024**

Recebemos dois arquivos Excel no formato .xlsb, após análise inicial, verificado que apresentaram as mesmas colunas e tipos de dados. Optei por unificar ambos em um único documento no Excel para centralizar as informações e facilitar o tratamento. Para a análise e tratamento dos dados, utilizamos Python pelas suas bibliotecas de análise de dados, como Pandas e NumPy. O arquivo foi convertido para o formato .xlsx para evitar problemas de compatibilidade.

**Etapa de Limpeza e Tratamento de Dados**

O primeiro passo foi carregar os dados unificados, garantindo que a coluna **CEP** permanecesse como string, preservando os zeros à esquerda. Criamos uma cópia do DataFrame para evitar alterações no original e identificamos a porcentagem de valores nulos por coluna, com as seguintes decisões:

* **HORA\_OCORRENCIA e DESC\_PERIODO:**
  + Removi o formato HMS.0000000 de **HORA\_OCORRENCIA** e ajustei para o formato de horas.
  + Com base nos horários, realizei o preenchimento da coluna **DESC\_PERIODO**: Madrugada (00:00-05:59), Manhã (06:00-11:59), Tarde (12:00-17:59), Noite (18:00-23:59) e Hora Incerta (valores indefinidos).
  + Preenchi a coluna **HORA\_OCORRENCIA** com valores padrão dependendo do período: Manhã (09:00), Tarde (15:00), Noite (22:00), Madrugada (03:00) e NULL para casos indefinidos.
* **Colunas com alta porcentagem de valores nulos (Acima de 50%):**
  + **TIPO\_INTOLERANCIA, DESDOBRAMENTO, CIRCUNSTANCIA e DESCR\_UNIDADE** foram removidas devido à baixa relevância e alta quantidade de valores nulos.
* **Colunas com até 5% de valores nulos:**
  + As linhas com valores nulos em **HORA\_OCORRENCIA, DATAHORA\_IMPRESSAO\_BO, DESCR\_TIPOLOCAL, DESCR\_SUBTIPOLOCAL, BAIRRO e MARCA\_OBJETO** foram removidas, resultando em uma perda tolerável na base de dados.
* **Colunas com mais de 5% de valores nulos:**
  + **DESC\_CONDUTA:** Preenchida com "Não especificado" para evitar distorções nos resultados.
  + **CEP:** Substituído pelo valor padrão "999999999".
  + **NUMERO\_LOGRADOURO:** Preenchida com o valor padrão "10".
  + **LATITUDE e LONGITUDE:** Preenchidas pela média dos valores por cidade.
  + **FLAG\_BLOQUEIO e FLAG\_DESBLOQUEIO:** Preenchidas pela moda.

Após o tratamento, eliminamos registros duplicados e salvamos o DataFrame no formato Excel para importação no Power BI.

**Modelagem e Análise Exploratória no Power BI**

Os dados foram importados para o Power BI, onde as colunas foram padronizadas de acordo com seus tipos de dados (texto, número, datas e hora). Criei duas dimensões com o objetivo de evitar linhas duplicadas nas análises distorcendo o resultado em análise de valores totais: **Dimensão boletim**, contendo informações relacionadas ao boletim de ocorrência, e **Dimensão bjeto**, com dados sobre o produto roubado. A tabela fato **Ocorrência** foi mantida para centralizar as análises.

Os principais indicadores gerados incluem o total de ocorrências entre janeiro e setembro de 2024, total de celulares furtados e a taxa de flagrante. Utilizamos gráficos para identificar padrões, destacando:

* **Principais marcas furtadas:** Samsung, Apple, Motorola e Xiaomi lideram as ocorrências.
* **Cidades mais afetadas:** As maiores incidências de furtos estão concentradas nas regiões centrais das cidades.
* **Período do dia com maior risco:** A noite concentra o maior número de ocorrências, seguida pela tarde.
* **Locais críticos:** A maioria dos furtos ocorre em vias públicas.
* **Evolução temporal:** Os casos aumentaram gradualmente de janeiro a setembro de 2024.

Além disso, um mapa de calor foi criado para representar a distribuição geográfica dos furtos no estado de São Paulo. Esse mapa destacou a concentração de ocorrências nas áreas urbanas centrais.

**Insights Principais**

A análise revelou que apenas 0,5% dos casos resultaram em flagrante, isso indica baixa efetividade das ações preventivas no momento do crime. A região central das cidades apresentou os maiores índices de ocorrências, provavelmente devido à maior densidade populacional e fluxo de pessoas. As vias públicas foram o local mais frequente das ocorrências, especialmente durante a noite, sugerindo maior vulnerabilidade nesse período. Por fim, as marcas Samsung, Apple, Motorola e Xiaomi foram as mais furtadas, provavelmente devido à popularidade e valor de venda dos dispositivos.