

Sprint 4

O SPC nos forneceu novos dados para trabalharmos. Os indicadores deste tratamento de dados levaram em consideração as quantidades de parcelas não pagas e a quantidade de dias de atraso das parcelas.

Para a quantidade de parcelas não pagas exigiu-se a filtragem das colunas “qtd_pcl_vnc” e “qtd_pcl_pgr” do arquivo “fatec_mvt” para campos não nulos e valores distintos dos nomes das próprias colunas (Fig1). As colunas armazenam os valores da quantidade de parcelas a vencer e quantidade de parcelas a pagar, conforme o dicionário do SPC respectivamente (Fig.2).

Figura 1: **A:** método *dropna()* da biblioteca *pandas* para a remoção de valores nulos e **B:** filtragem dos valores para ficar com as linhas dos campos referente a dados diferentes dos nomes das colunas.

```
A d = c.dropna(subset=['qtd_pcl_vnc'])
    d = c.dropna(subset=['qtd_pcl_pgr'])
    d

B d = d[d['qtd_pcl_vnc'] != d['qtd_pcl_pgr']]
    d
```

Figura 2: Dicionário do SPC. Coluna 1: nome do arquivo; coluna 2: nome da coluna; coluna 3: significado do nome da coluna.

mvt	qtd_pcl_vnc	quantidade de parcelas a vencer
mvt	qtd_pcl_pgr	quantidade de parcelas a pagar

Após a filtragem dos dados, necessitou-se alterar os tipos de dados de *string* para números inteiros. Esse tipo inicial de dados foi devido à existência de nomes das colunas e valores nulos nos campos (Tab. 1).

Tabela 1: Exibição dos nomes das colunas à esquerda e dos tipos de dados armazenados à direita. Out[165]: o tipo *object* é uma *string* ou caracter. Nota-se que nas colunas “qtd_pcl_vnc” e “qtd_pcl_pgr” os valores não são números inteiros. In[166]: para a conversão para números inteiros, aplicou-se o método `astype()`.

```
In [165]: d.dtypes
Out[165]: nulo          object
          id_opr_cad_pos  object
          dat_vct         int32
          qtd_pcl_vnc     object
          qtd_pcl_pgr     object
          vlr_tot_fat_tfm  object
          vlr_min_fat_tfm  object
          vlr_pcl_tfm     object
          tip_mvt         object
          prd             object
          dtype: object

In [166]: d['qtd_pcl_vnc'] = d['qtd_pcl_vnc'].astype('int')
          d['qtd_pcl_pgr'] = d['qtd_pcl_pgr'].astype('int')
```

A quantidade de parcelas em atraso não está evidenciada na tabela. Para descobri-las, fez-se uma função que subtrai a coluna “qtd_pcl_pgr” da coluna “qtd_pcl_vnc” (parcelas a pagar menos parcelas a vencer), e o resultado, que é a quantidade de parcelas vencidas, foi armazenado na coluna “nulo” (Fig.4). Os dados foram coletados e plotados em um gráfico de barras horizontais. Os valores à esquerda mostram quantas parcelas atrasadas; à direita, é exibido a quantidade de pessoas para as referidas parcelas atrasadas (Tab. 2 e Fig.5).

Figura 4: interpretação da função `sub()` e sua aplicação na coluna “nulo”. In[169]: a função subtrai a quantidade de parcelas a pagar da quantidade de parcelas a vencer; In[179]: uso do método `apply()` com o nome da função entre parênteses, com o retorno na coluna “nulo”.

```
In [169]: def sub(g):  
           g = d['qtd_pcl_pgr'] - d['qtd_pcl_vnc']  
           return g  
  
In [179]: d['nulo'] = d.apply(sub)
```

A quantidade de dias atrasados foi obtida no arquivo “fatec_pgt”. Foi feito a mesma operação do arquivo anterior de remoção de valores nulos com o método `dropna()` e remoção de valores iguais ao nome da própria tabela. Destaca-se aqui o uso do método `pd.to_datetime('nome da coluna')` que é um método para mudança de tipos dos campos da coluna referenciada (Fig. 6).

As datas foram formatadas usando o código de valores nulos e o código data, documentado na entrega 1, para a posterior aplicação dos métodos supracitados. A codificação de uma função de subtração foi formulada para resolver a operação aritmética das datas com o seu armazenamento na coluna “nulo” (Fig. 7). Obteve-se resultado zero negativos e maiores que zero. Os resultados zero são pagamentos feitos no prazo; negativos são pagamentos feitos adiantados e maiores que zero são pagamentos atrasados (Tab. 3). Filtrou-se somente os pagamentos atrasados e montou-se um gráfico de barras (Fig. 8); observou-se a ocorrência de nove picos relevantes. Obteve-se seus valores com a definição de limites no cálculo 1.

Para saber os valores exatos dos picos em destaque, restringiu-se os valores “x” e “y” do plano cartesiano para achar uma coordenada de maior probabilidade de seus valores, ou seja, quando Δx e Δy tendem à coordenada dos picos. Foi-se reduzindo o escopo cada vez mais até ser possível montar o gráfico da figura 9.

Tabela 2 e Figura 5: Tabela e gráfico das parcelas em atraso junto com a quantidade de pessoas para cada atraso. Na tabela conta-se a quantidade de pessoas (à direita) com suas referidas parcelas em atraso (à esquerda); O gráfico da tabela é exibido logo abaixo.

```
In [184]: d = d['nulo'].sort_values()
          d.value_counts()

Out[184]: 1      172
          2       66
          3        8
          4         3
          10        2
          82         1
          14         1
           9         1
           8         1
           6         1
          Name: nulo, dtype: int64
```

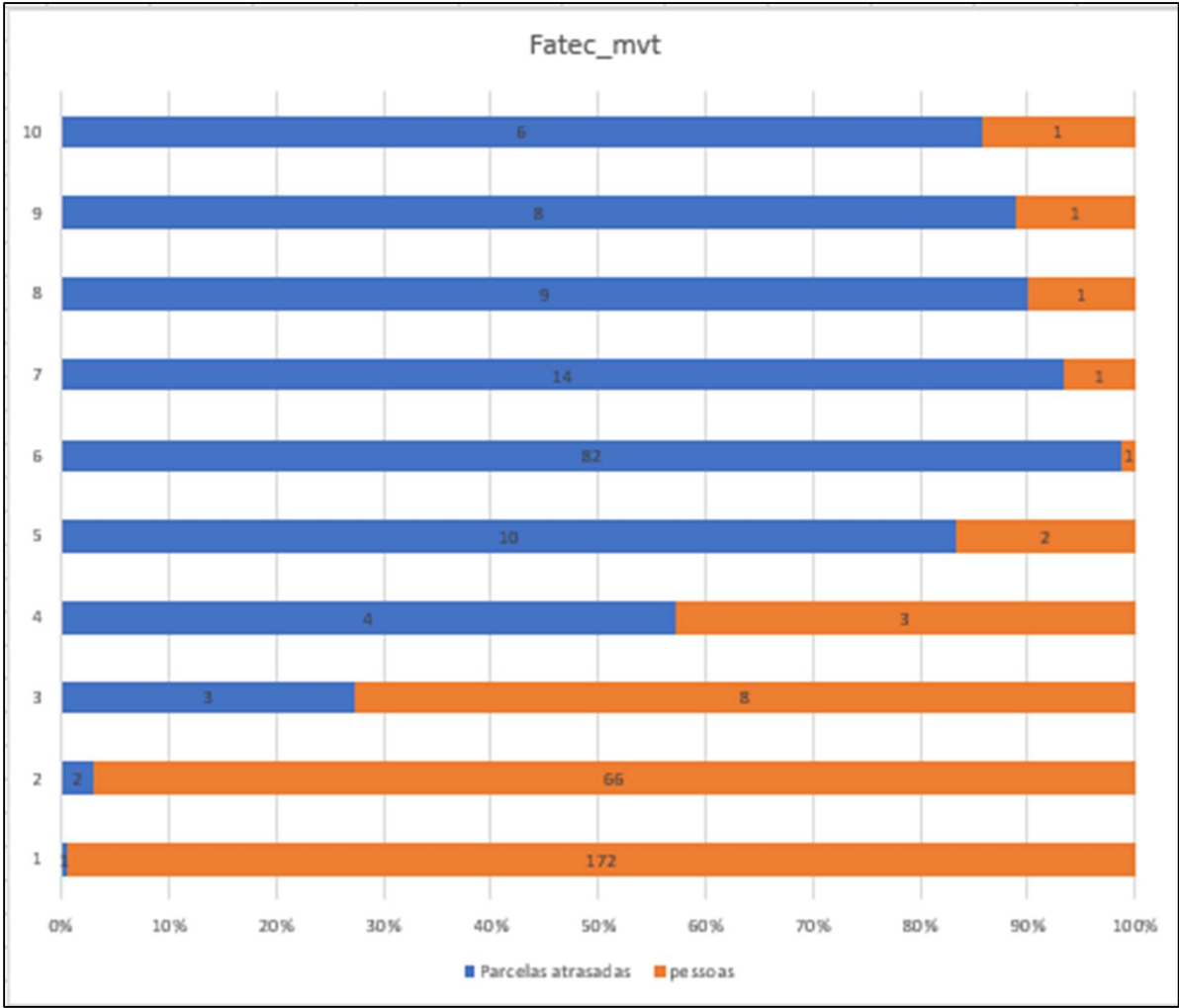


Figura 6: O código mostra a mudança dos campos da coluna “dat_pgt” para o formato data.

```
In [46]: a['dat_pgt'] = pd.to_datetime(a['dat_pgt'])
a['dat_vct_tfm'] = pd.to_datetime(a['dat_vct_tfm'])
a.info()
```

Figura 7: fórmula de subtração das datas com o nome “atraso” e sua aplicação na coluna “nulo” com o método *apply()*.

```
In [88]: def atraso(g):
g = a['dat_pgt'] - a['dat_vct_tfm']
return g

In [106]: a['nulo'] = a.apply(atraso)
```

Tabela 3: Resultado da aplicação da função da figura 7. Os valores zero são pagamentos no prazo; negativos, adiantados e maiores que zero, atrasados.

nulo
0 days
-58 days
-518 days
-28 days
247 days
...
-2 days
-216 days
-59 days
-89 days
4 days

Figura 8: Gráfico mostrando dias atrasados em função de pessoas.

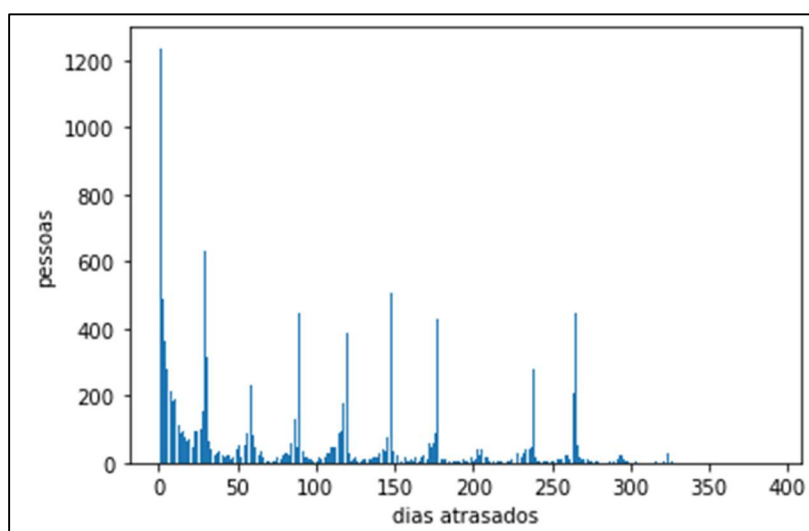
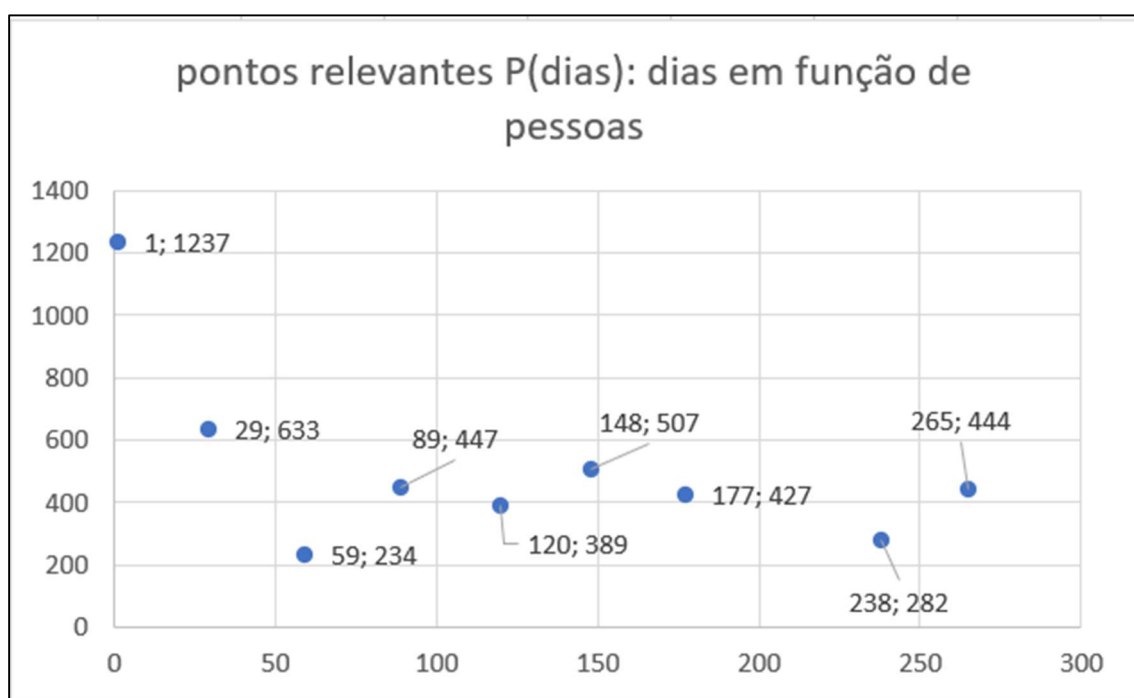


Figura 9: Gráfico evidenciando os nove picos que se destacaram da figura 8. Repare que existe uma média de 30 dias de distância entre picos, com exceção do antepenúltimo pico (177;427) e penúltimo pico (238;282), que tem uma distância aproximada de 60 dias.



Conforme visualizado na figura 9, houve uma tendência de atrasos mensais com exceção das coordenadas (177;427) e (238;282) com uma diferença de dois meses.