1.Business and Data Understanding

“O **training set** deverá ser usado para desenvolver e treinar o modelo de Machine Learning. No training set é fornecida informação referente ao nível de incidentes rodoviários (**accidents**) de cada registo para que possam treinar os modelos de aprendizagem. O modelo a desenvolver deverá ter, na sua base, features como a magnitude do atraso que se verifica numa determinada hora, o tempo de atraso provocado pelos incidentes, a temperatura, pressão atmosférica e a velocidade do vento, entre outras features que caracterizam um determinado ponto temporal.”

“O **test set** será utilizado para validar a performance do modelo em dados ainda não vistos pelo mesmo. Neste dataset de teste não é fornecida a class referente ao nível de incidentes rodoviários. O vosso trabalho é prever essa mesma class. Devem utilizar o modelo desenvolvido para prever, para cada registo do test set, o nível de incidentes rodoviários esperados.”

É também disponibilizado um exemplo de um ficheiro de submissão (example\_submission.csv) onde é assumido que o nível de incidentes rodoviários esperado para cada registo é Medium.

(EXPLICAR ISTO POR PALAVRAS NOSSAS)

# Features

city\_name - nome da cidade em causa;

record\_date - o timestamp associado ao registo;

magnitude\_of\_delay - magnitude do atraso provocado pelos incidentes que se verificam no record\_date correspondente;

delay\_in\_seconds - atraso, em segundos, provocado pelos incidentes que se verificam no record\_date correspondente;

affected\_roads - estradas afectadas pelos incidentes que se verificam no record\_date correspondente;

luminosity - o nível de luminosidade que se verificava na cidade de Braga;

avg\_temperature - valor médio da temperatura para o record\_date na cidade de Braga;

avg\_atm\_pressure - valor médio da pressão atmosférica para o record\_date na cidade de Braga;

avg\_humidity - valor médio da humidade para o record\_date na cidade de Braga;

avg\_wind\_speed - valor médio da velocidade do vento para o record\_date na cidade de Braga;

avg\_precipitation - valor médio de precipitação para o record\_date na cidade de Braga;

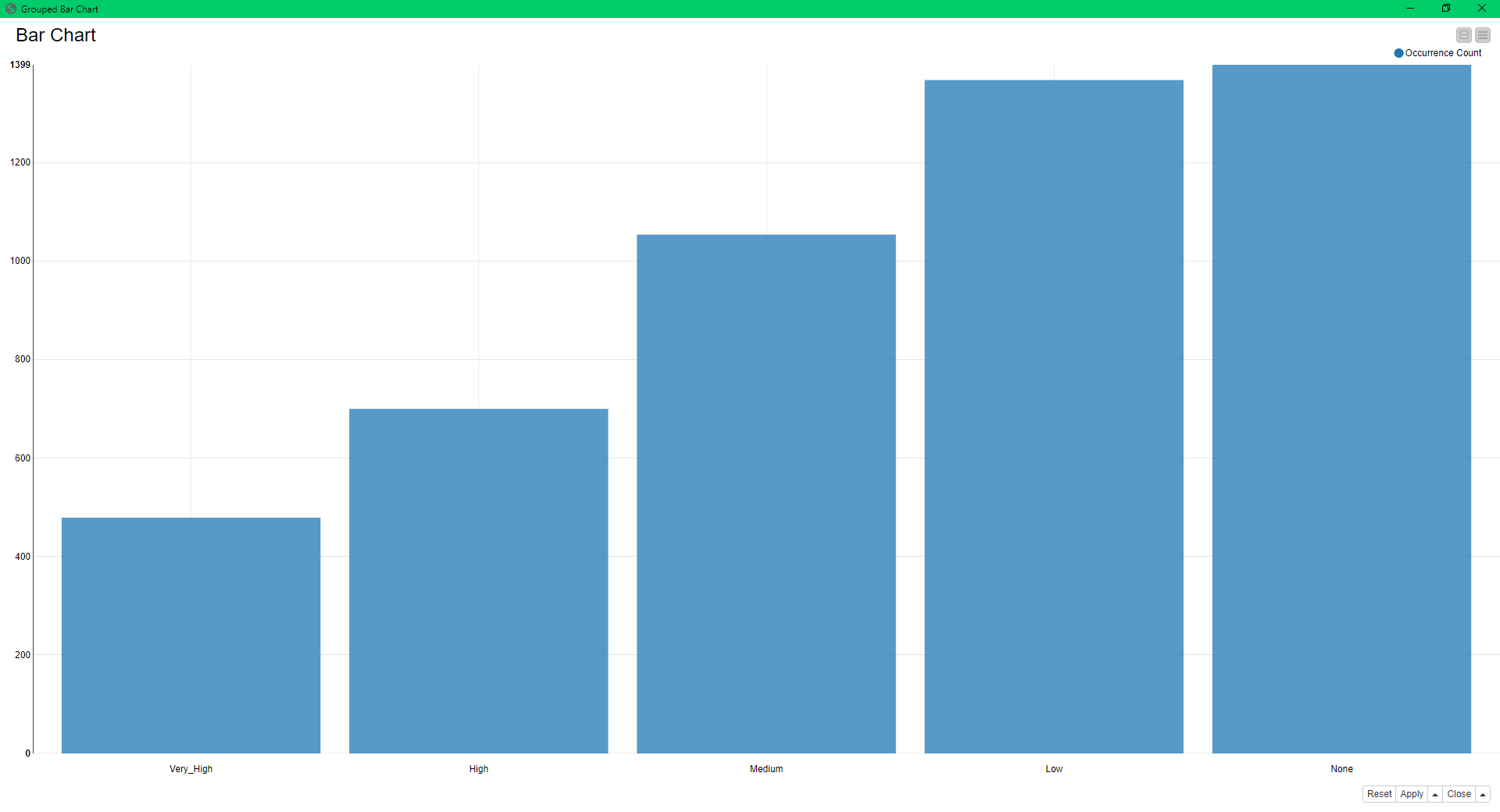
avg\_rain - avaliação qualitativa do nível de precipitação para o record\_date na cidade de Braga;

accidents - indicação acerca do nível de incidentes rodoviários que se verificam no record\_date correspondente na cidade de Braga.

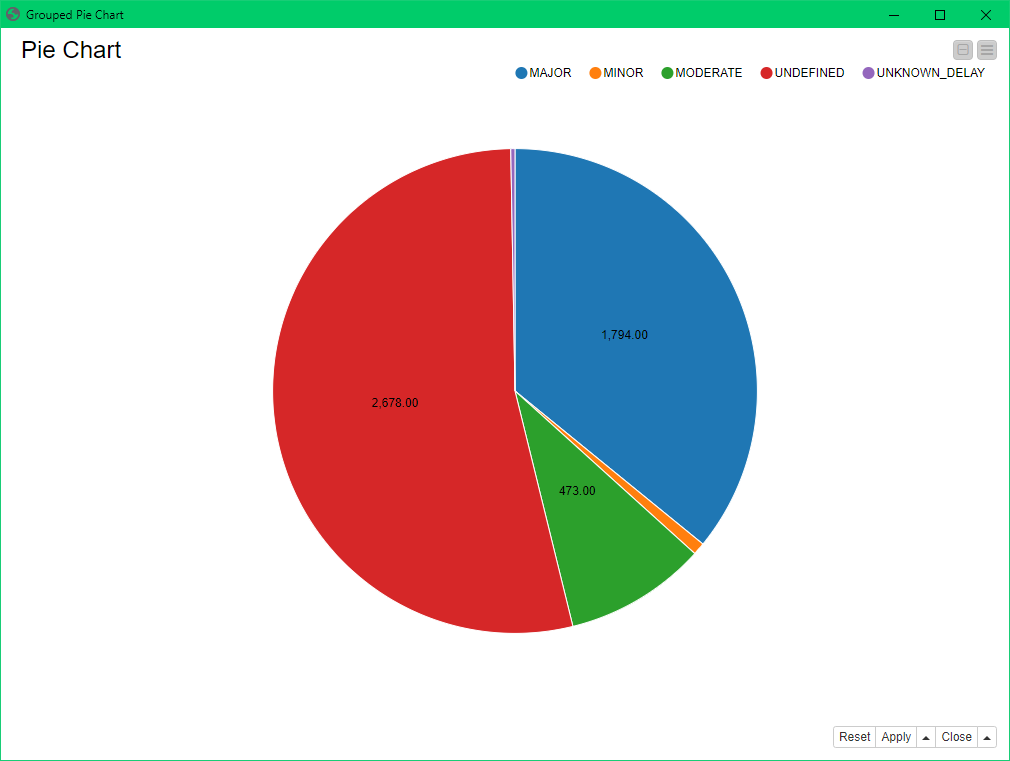
ETL FIRST COMPONENT

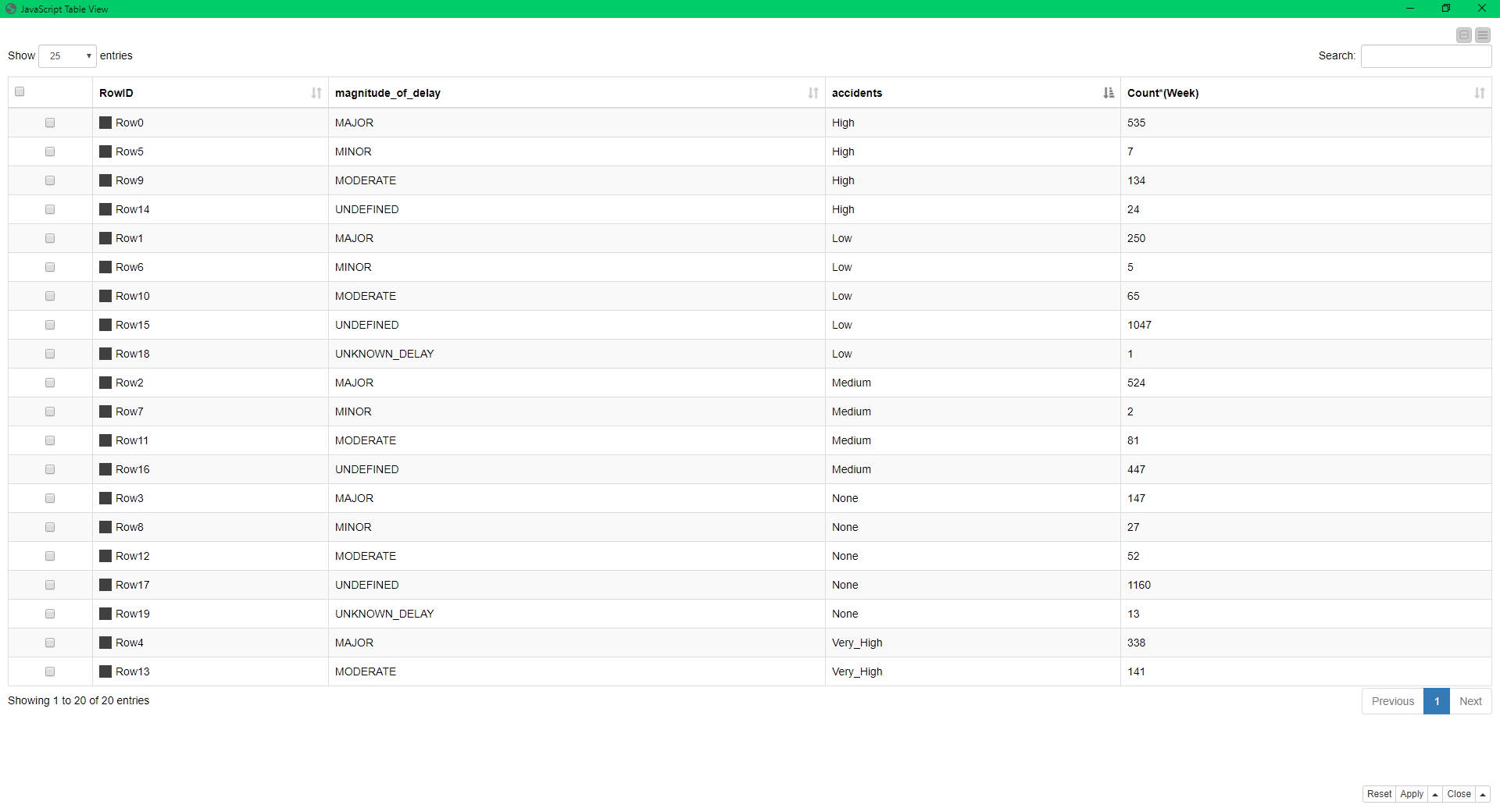
DATA ANALYSES

DISTRIBUIÇÃO DO NIVEL DE ACIDENTE

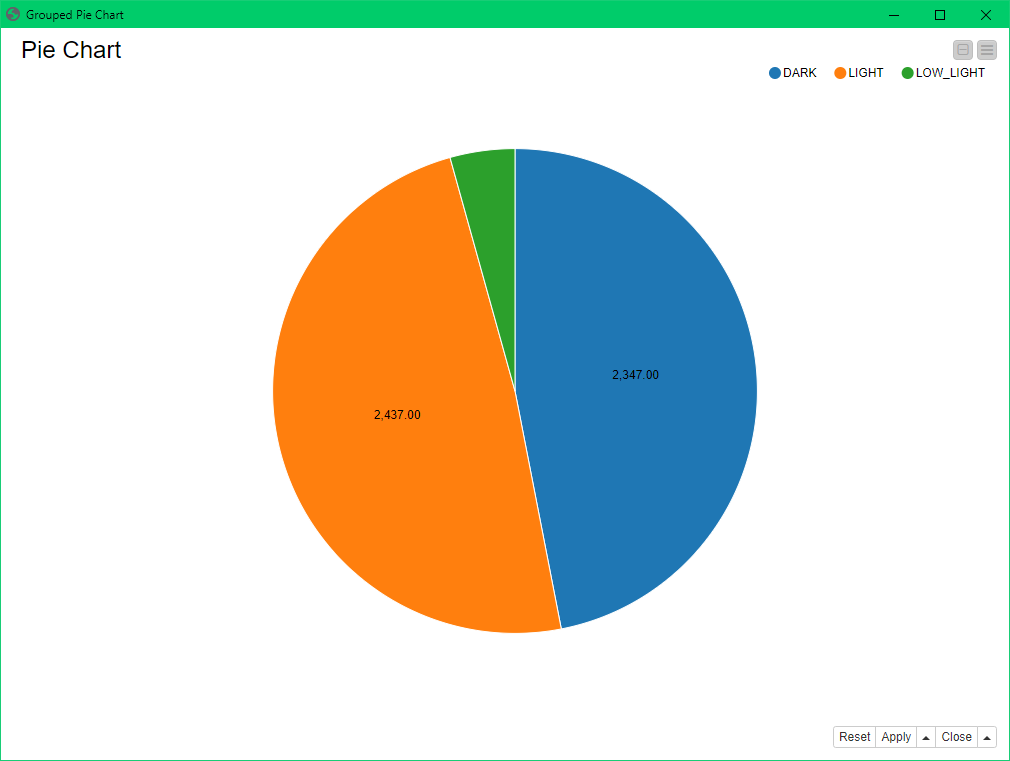


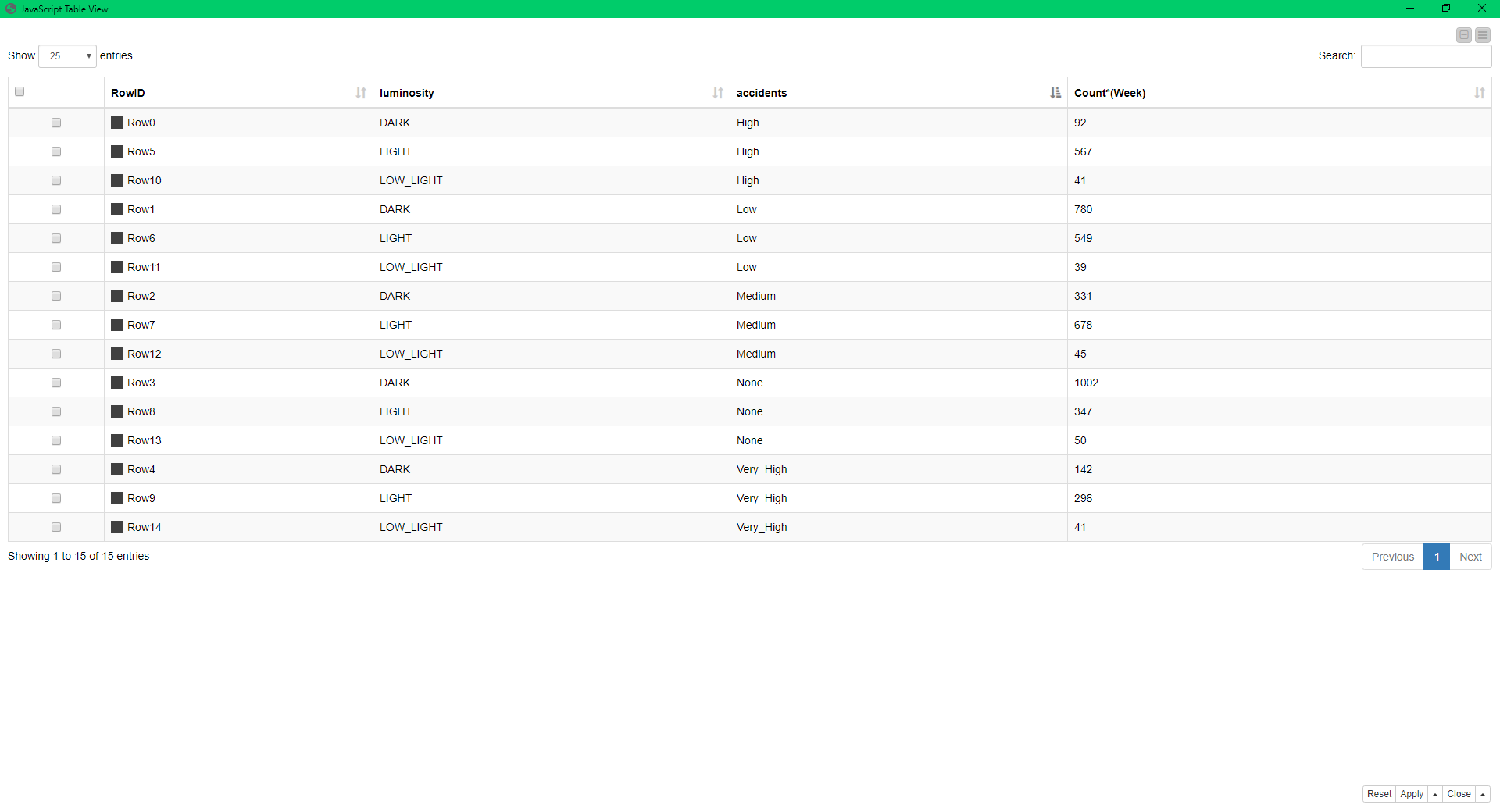
DISTRIBUIÇÃO DA MAGNITUDE OF DELAY



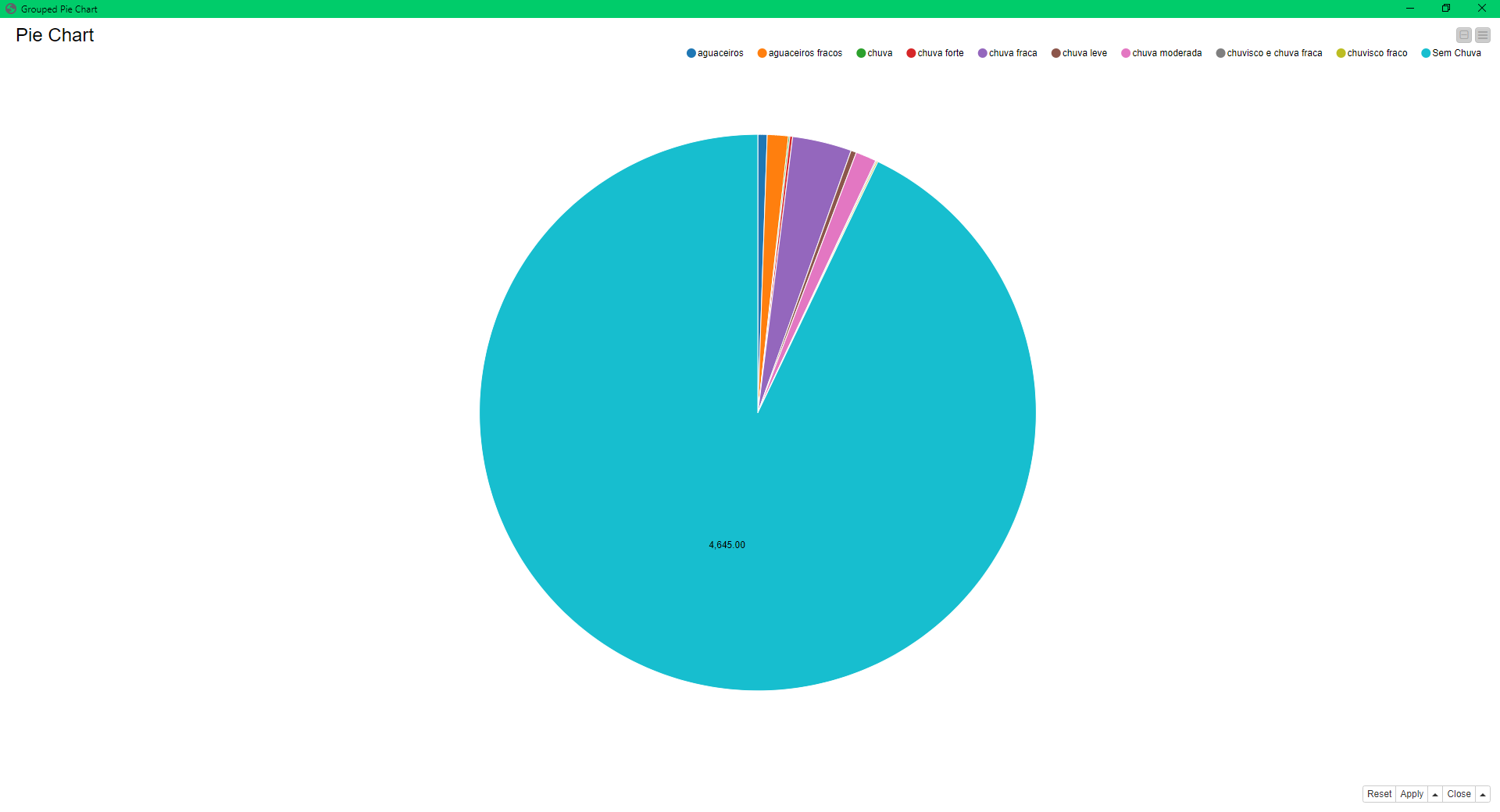


DISTRIBUIÇÃO DA LUMINUSIDADE (CONTAGEM DE ACIDENTES)





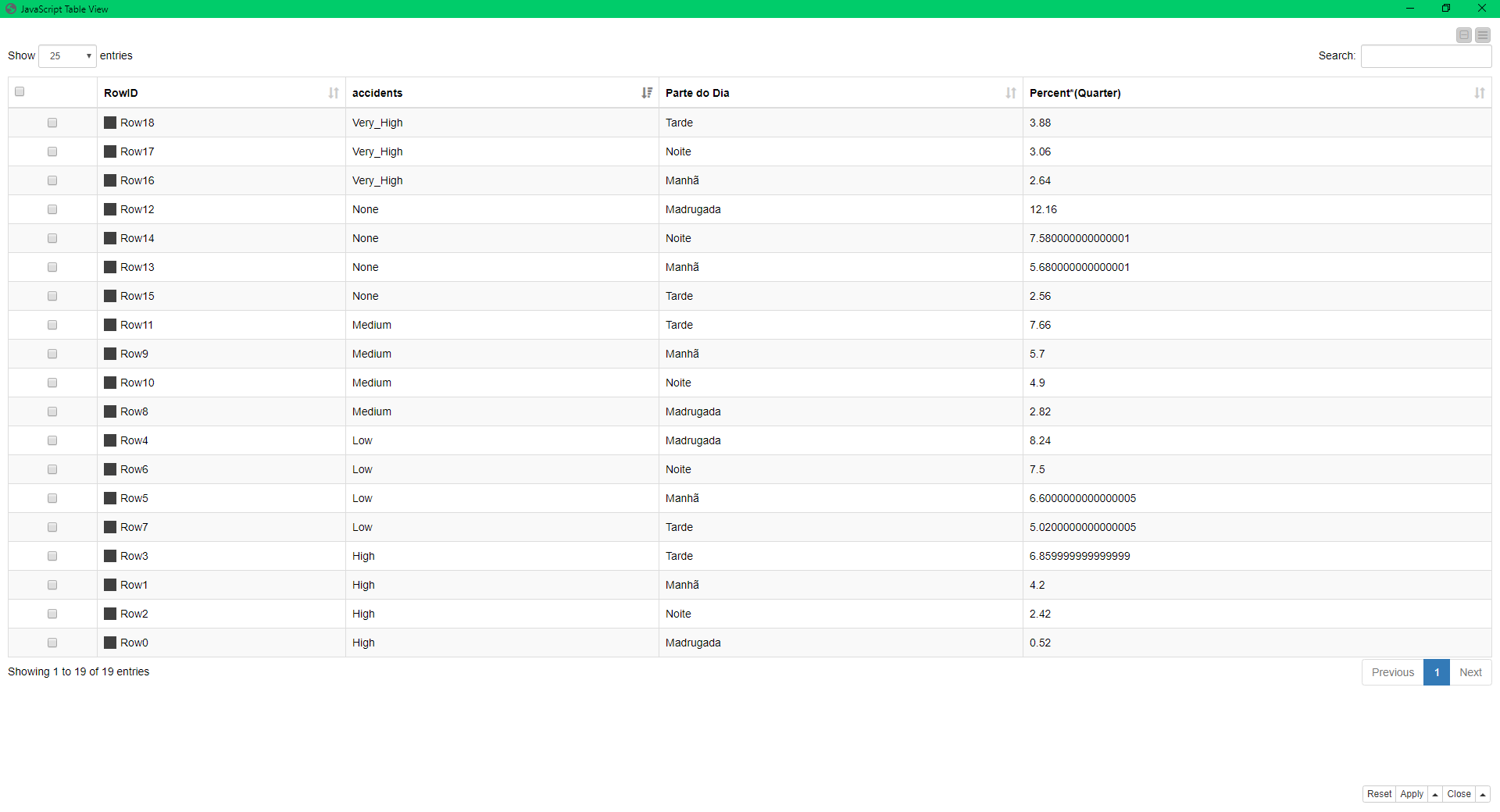
DISTRIBUIÇÃO DA CHUVA



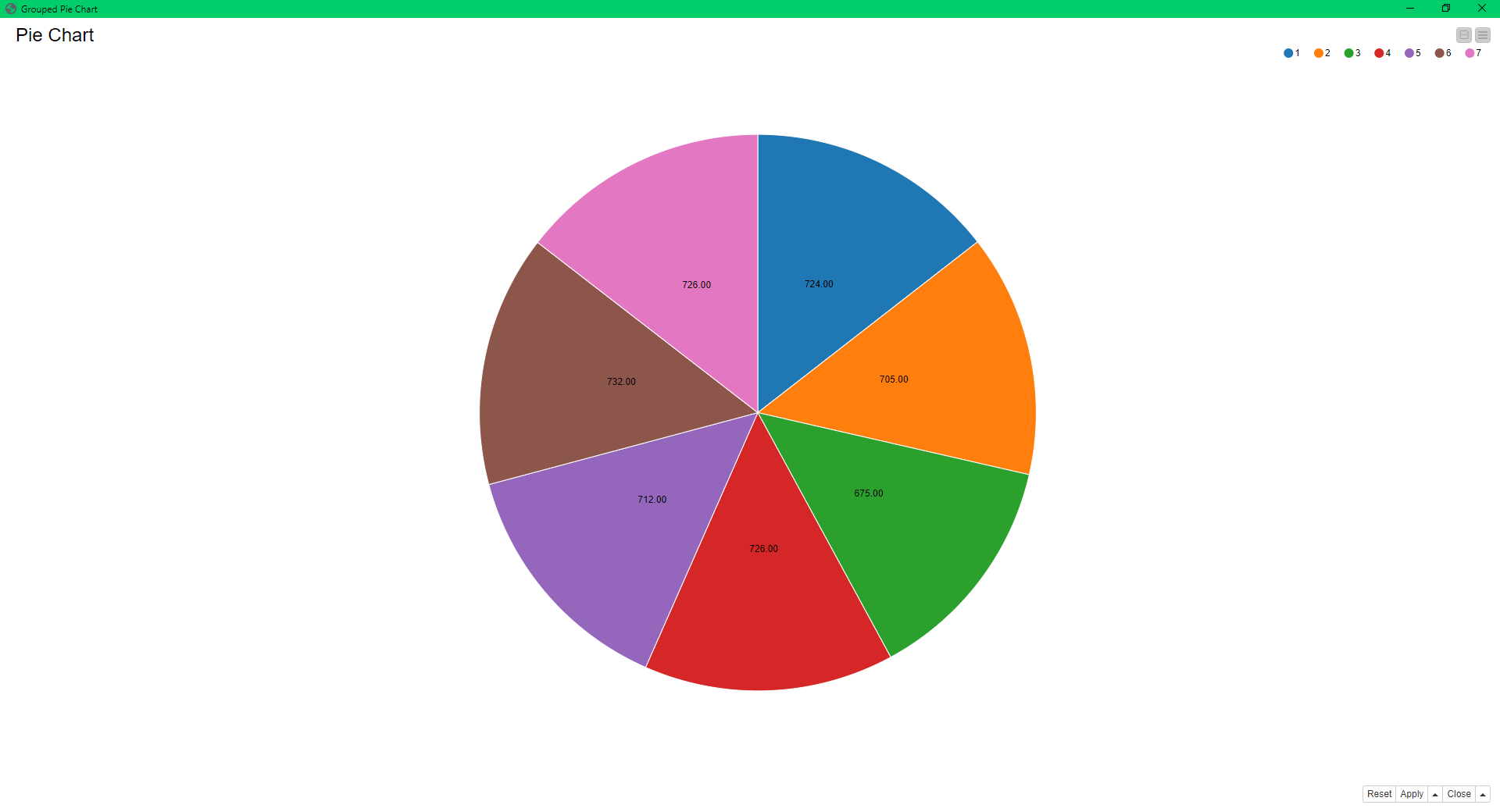
(Probabilidade de acidente grave para cada nível de chuva;)

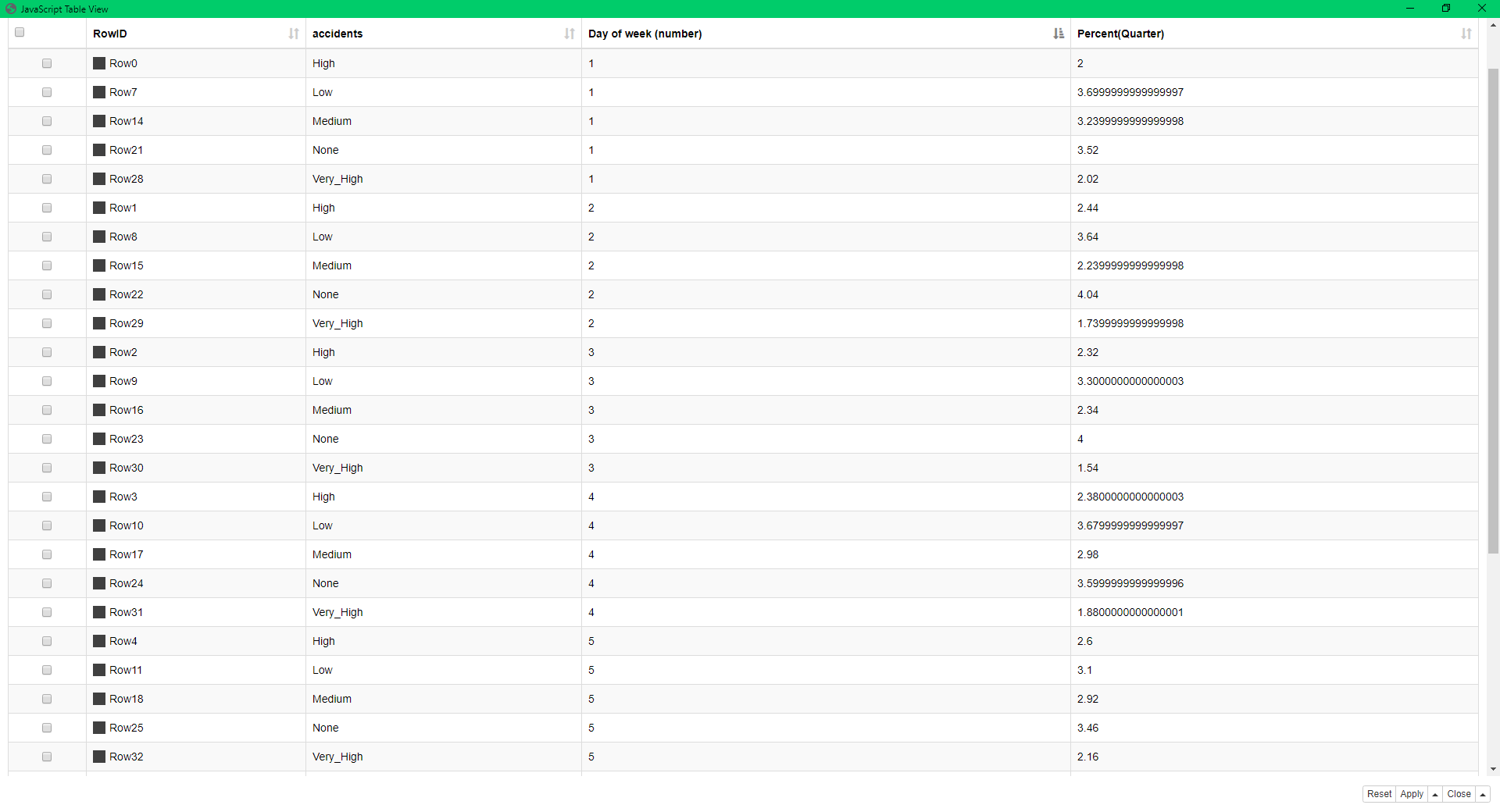
(Probabilidade de acidente com chuva por trimestre;)

PARTE DO DIA – ACIDENTES



DIA DA SEMANA

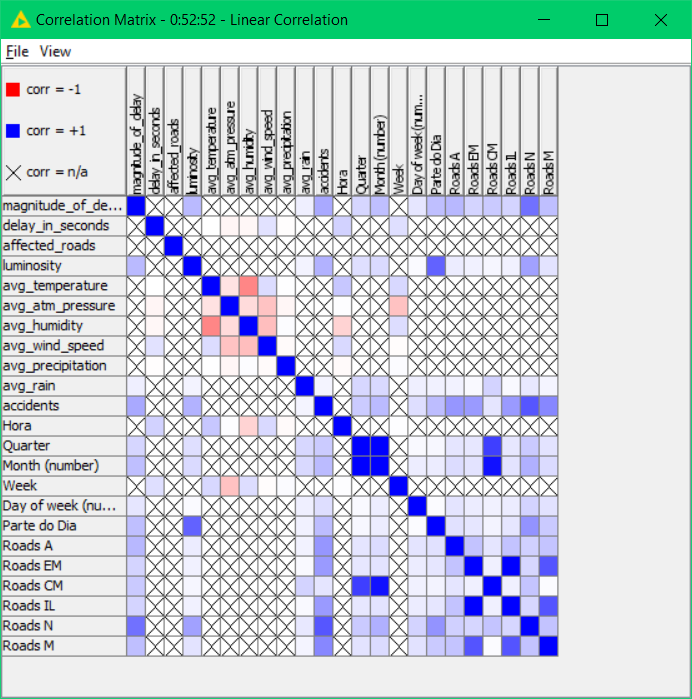






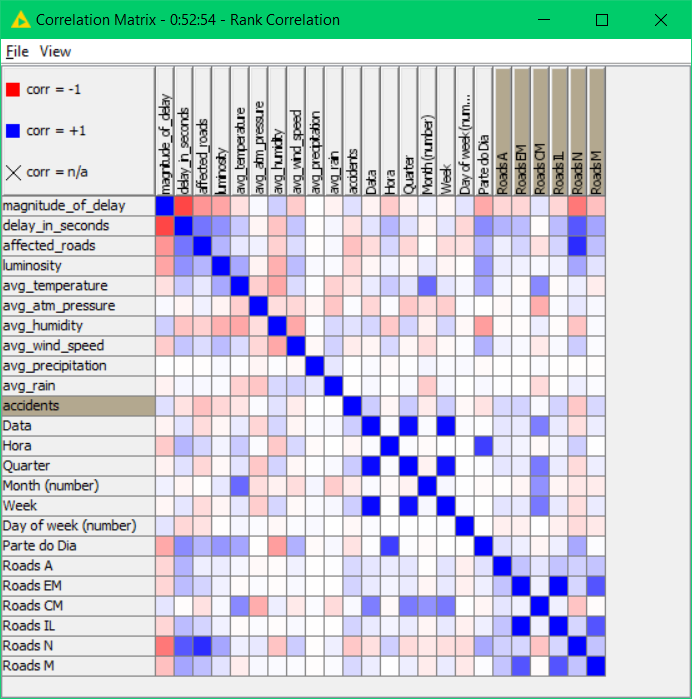
## Correlação Entre Todas as Variáveis

Linear

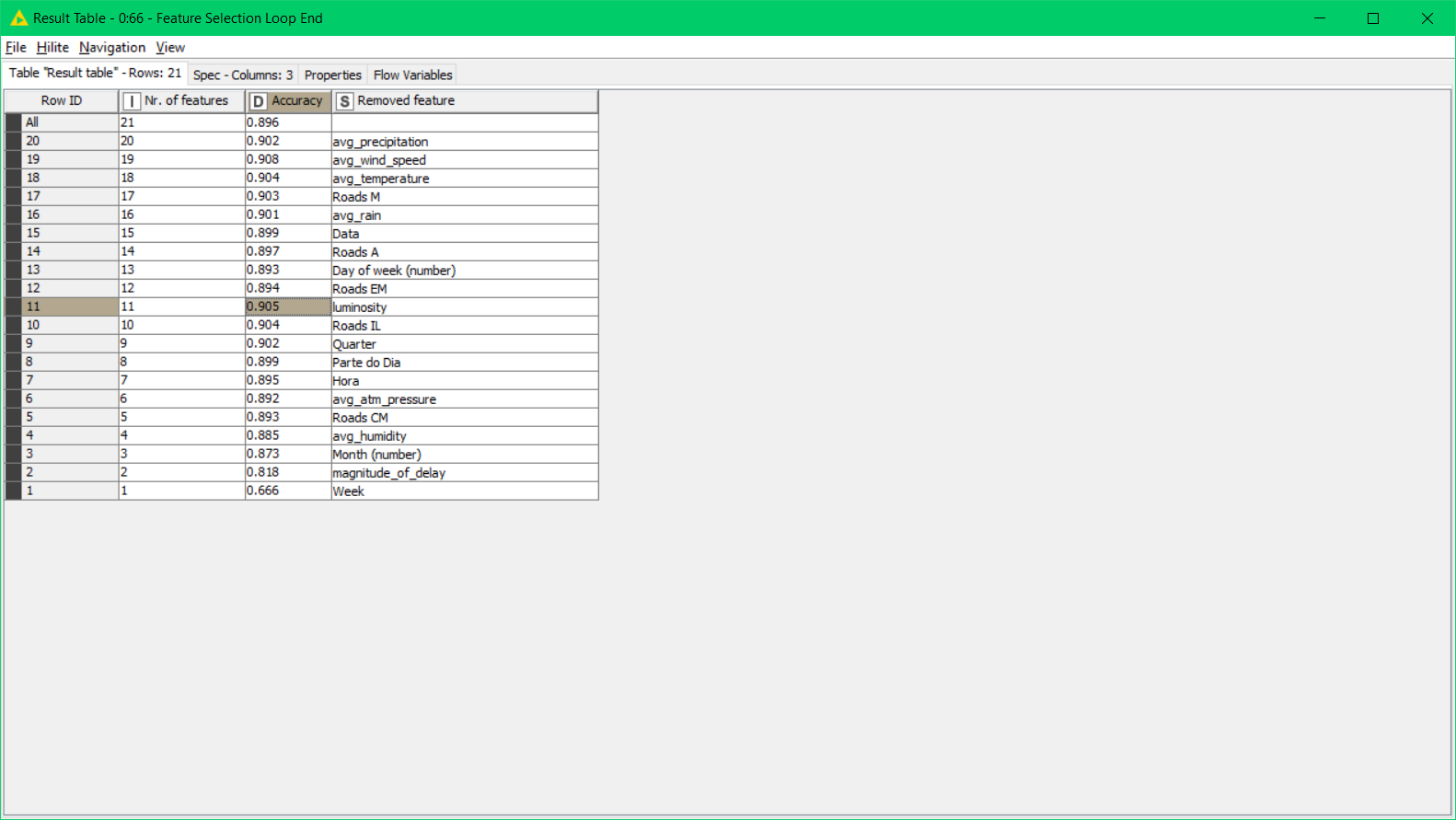


1. Como podemos ver a variável mais correlacionada com accidents é a Roads N;
2. A Magnitude\_of\_Delay e o Delay em segundos são muito correlacionados;
3. A Magnitude of Delay e o Roads N também são muito correlacionadas;
4. A luminosidade e a parte do dia estão muito correlacionadas;
5. A Avg\_temperature e a Parte do dia estão correlacionadas;
6. A Avg Precipitation não é correlacionada com nada no modelo mas também sabemos que há poucos registos de precipitação;
7. A Avg Rain tem correlação com algumas variáveis do modelo
8. Hora e parte do dia estão bastante correlacionadas
9. O quarter e o Month estão muito correlacionados com o Roads CM
10. Road CM com Road IL e M
11. Road IL com EM

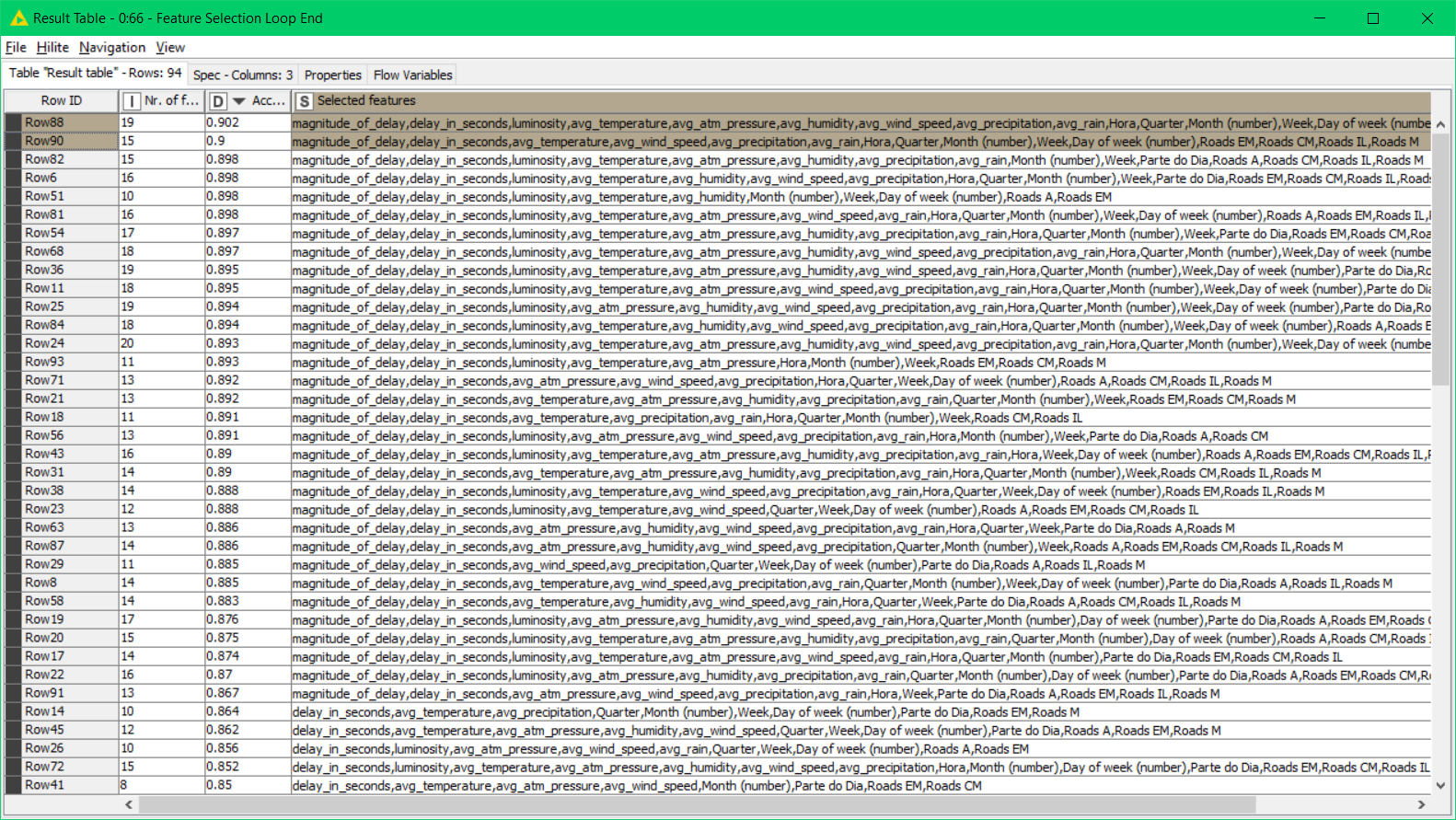
Spearman



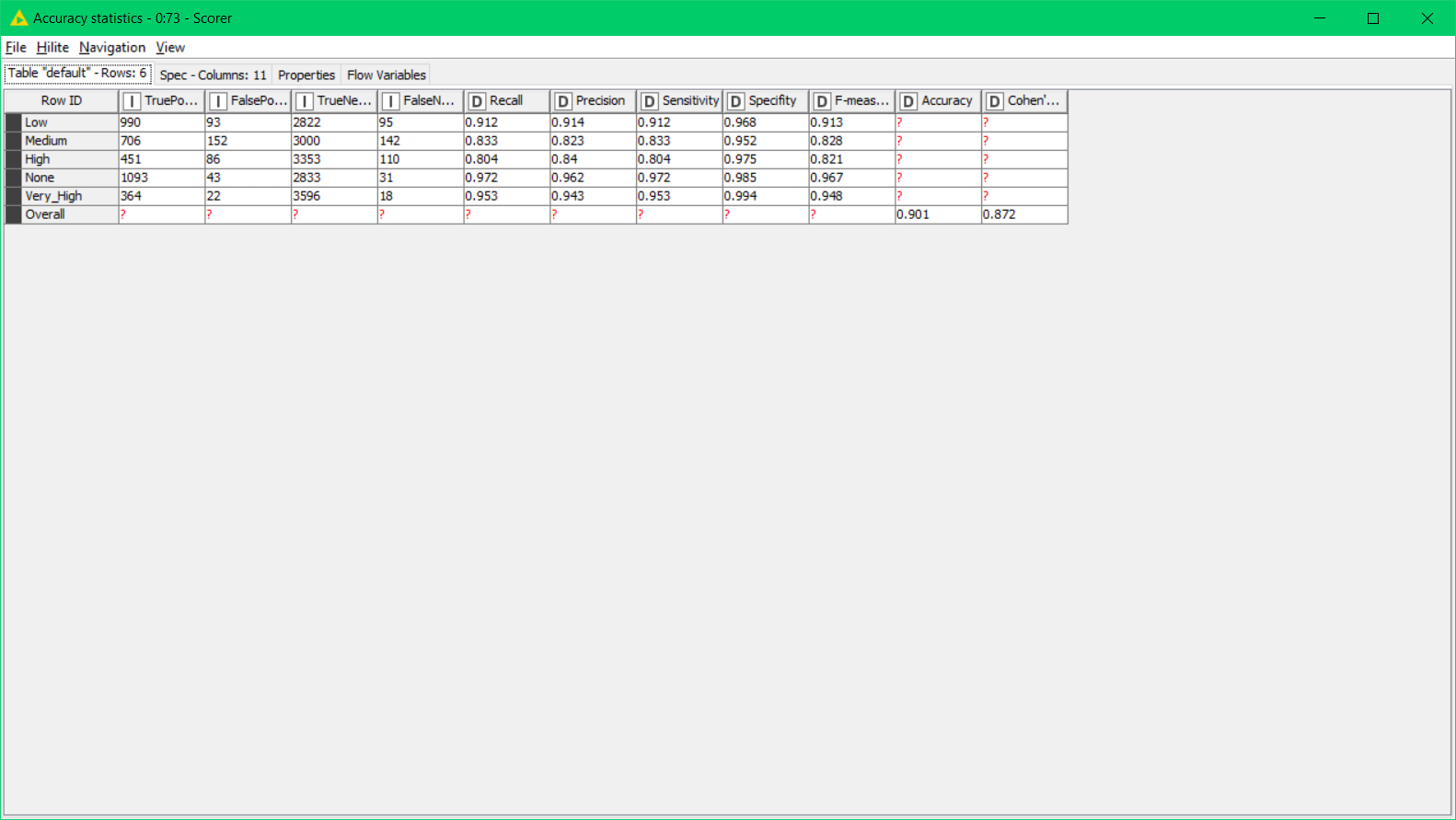
Best Labels : Backward Feature Elimination



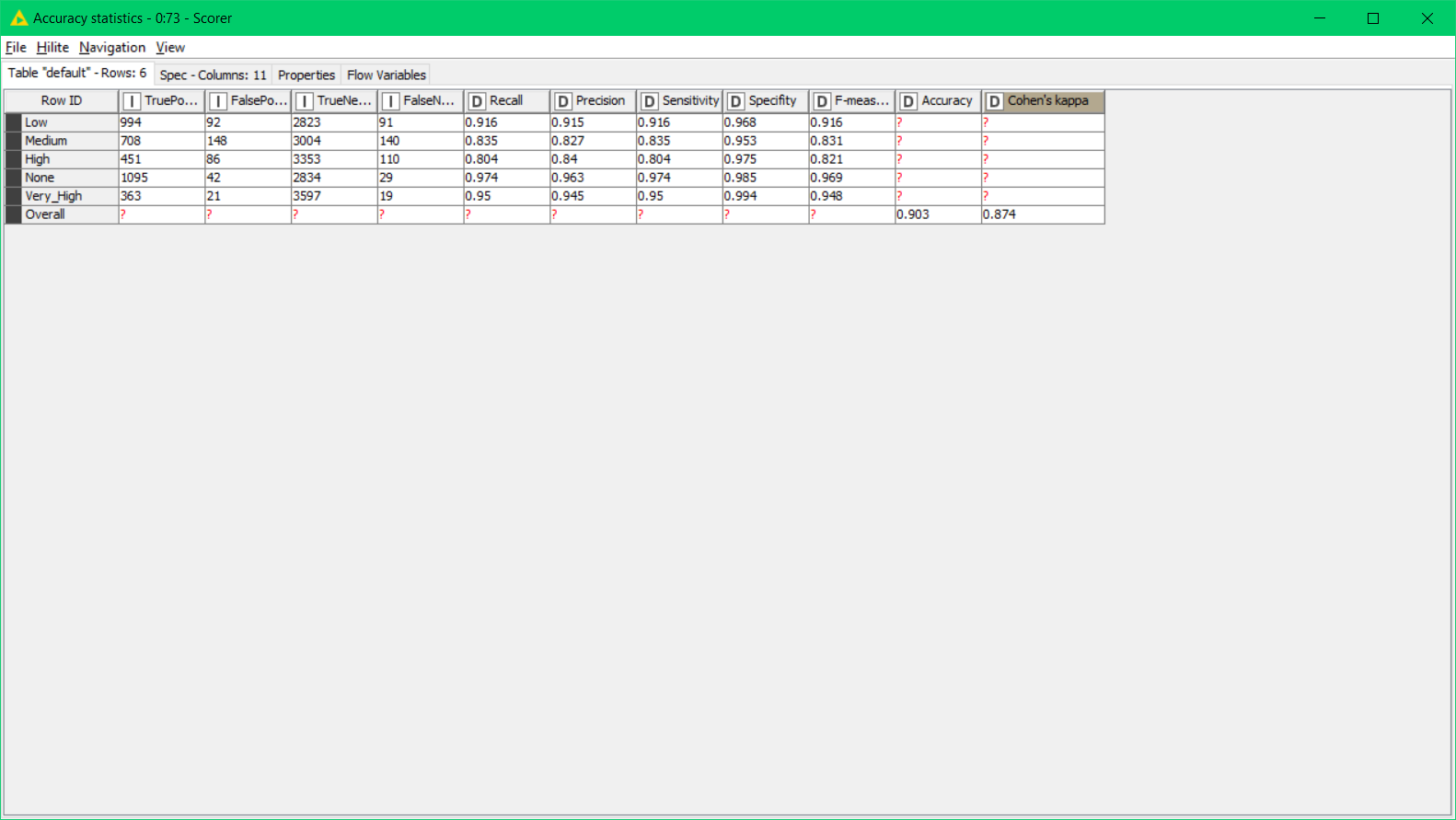
Best Labels Random



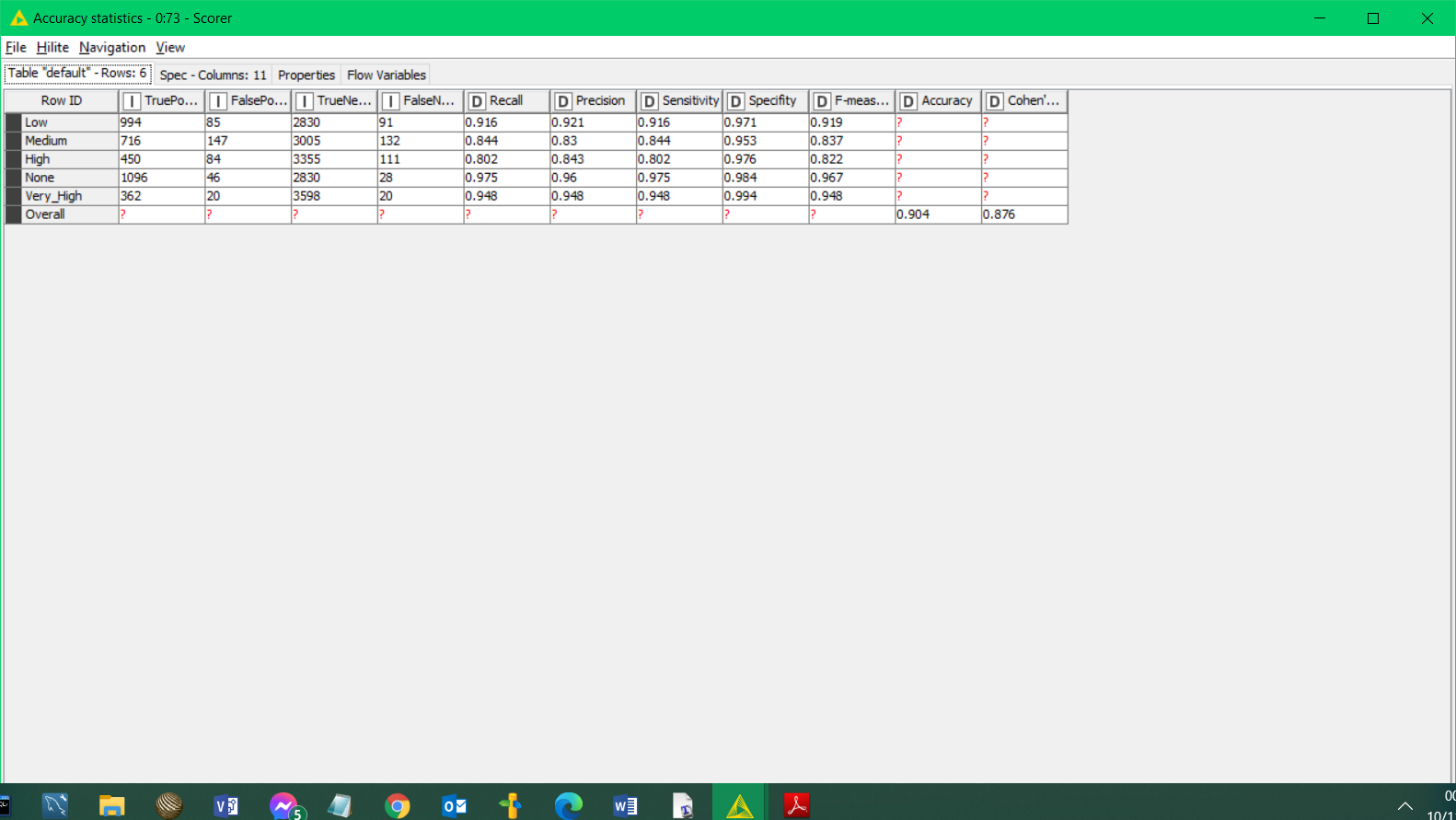
Retirar so avg precepitation



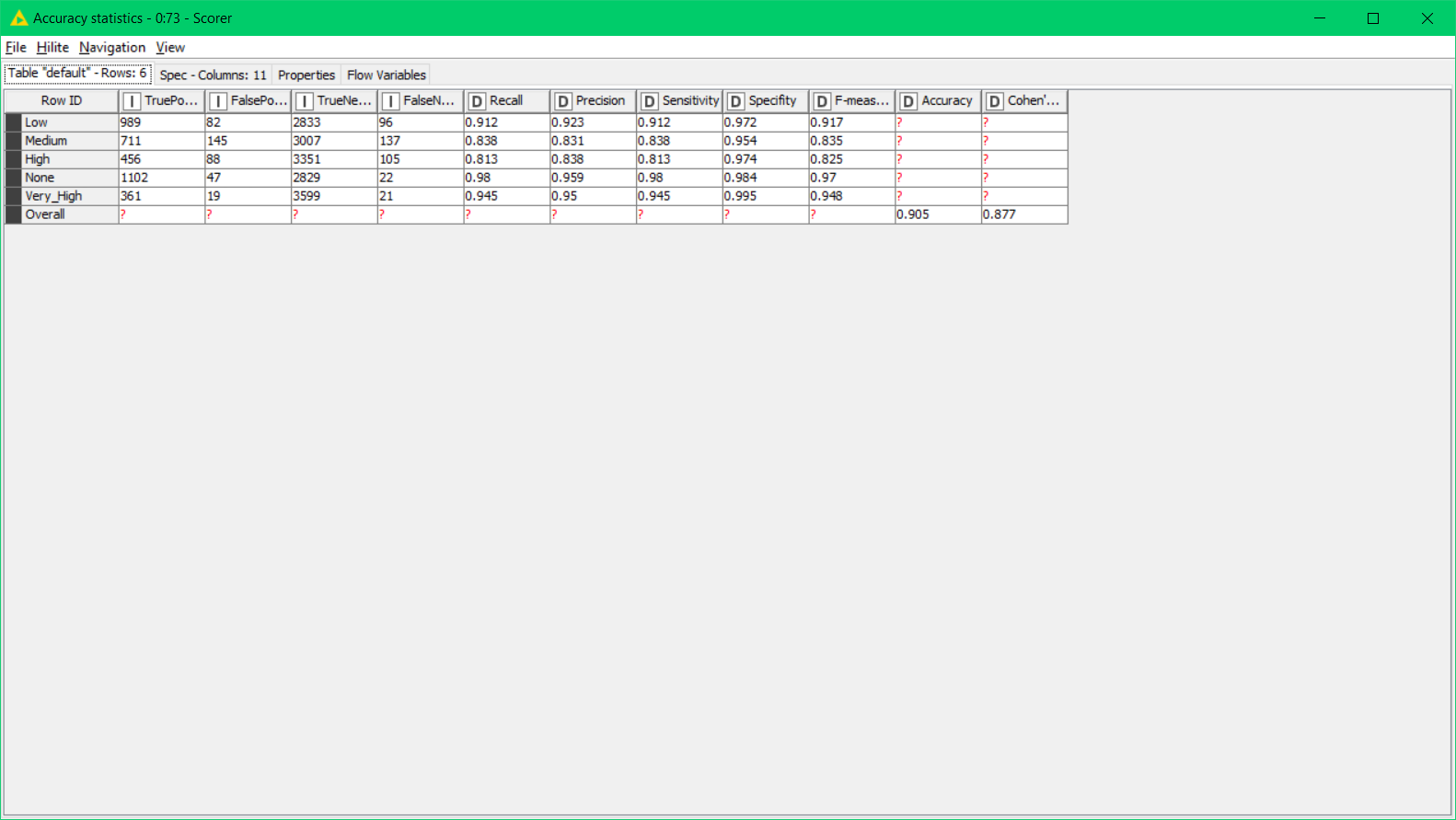
Retirar so avg precepitation e Roads M



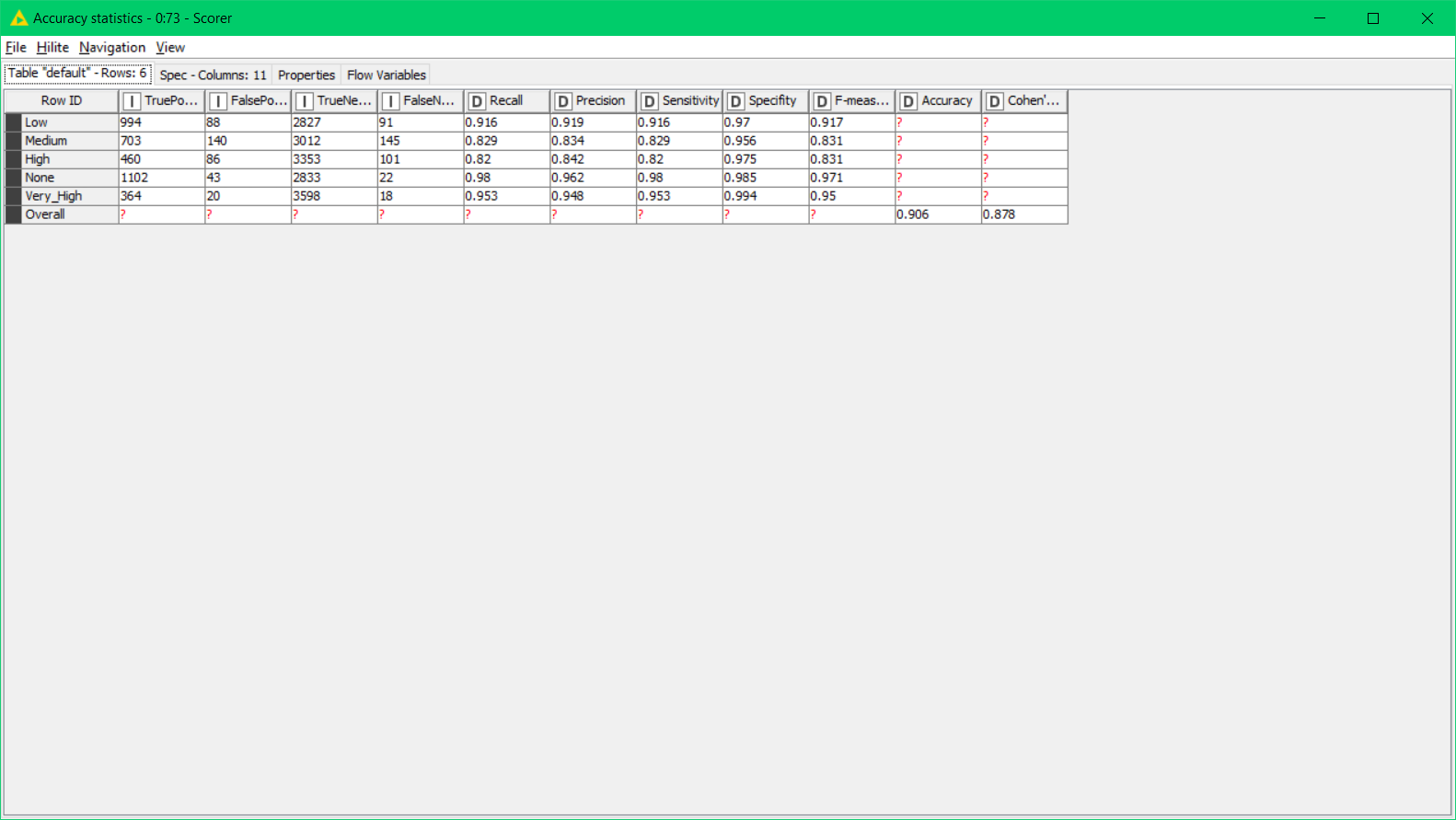
Retirar so avg precepitation e Roads M e Roads EM



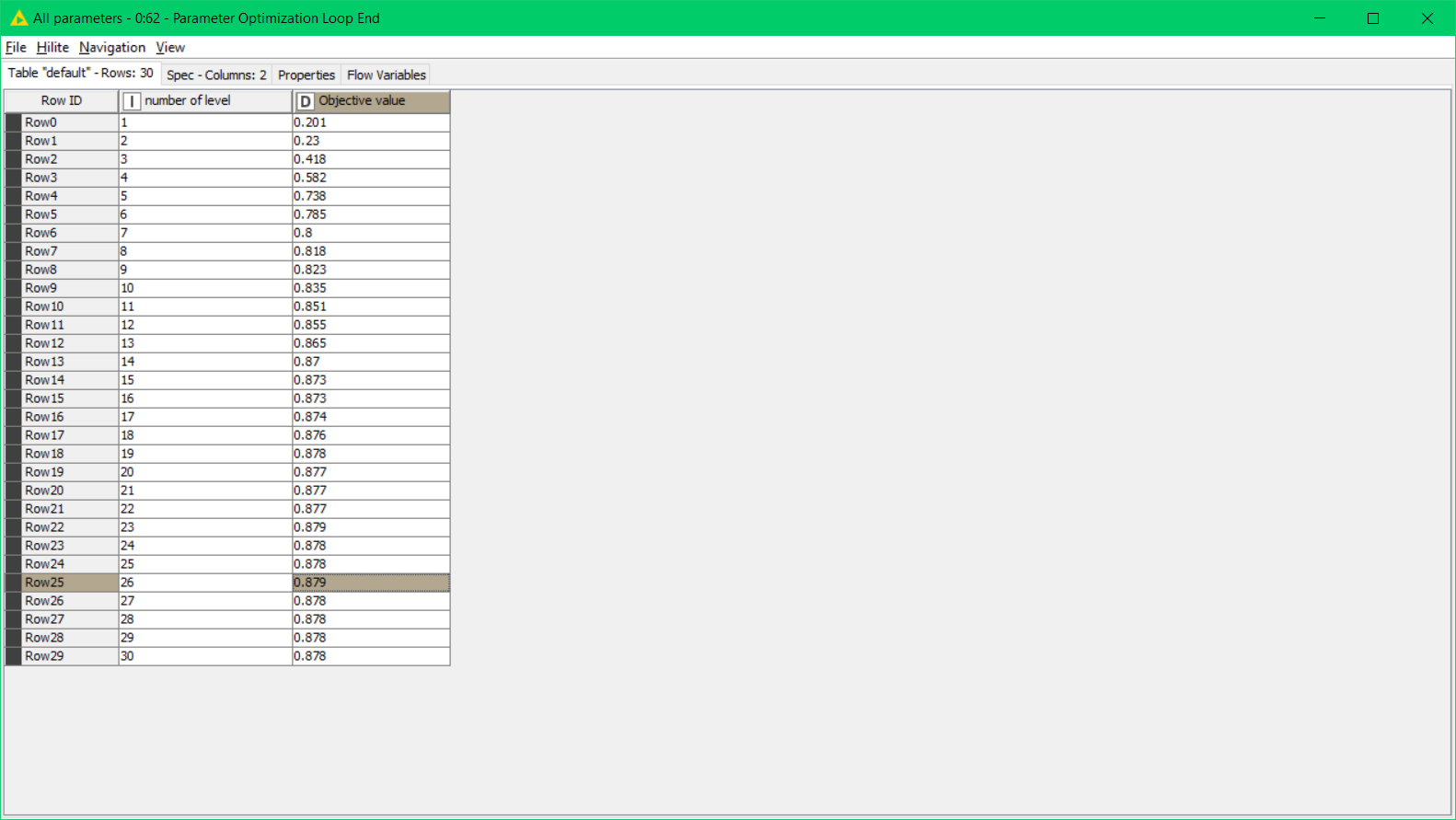
Retirar so avg precepitation e Roads M e Roads EM e Hora



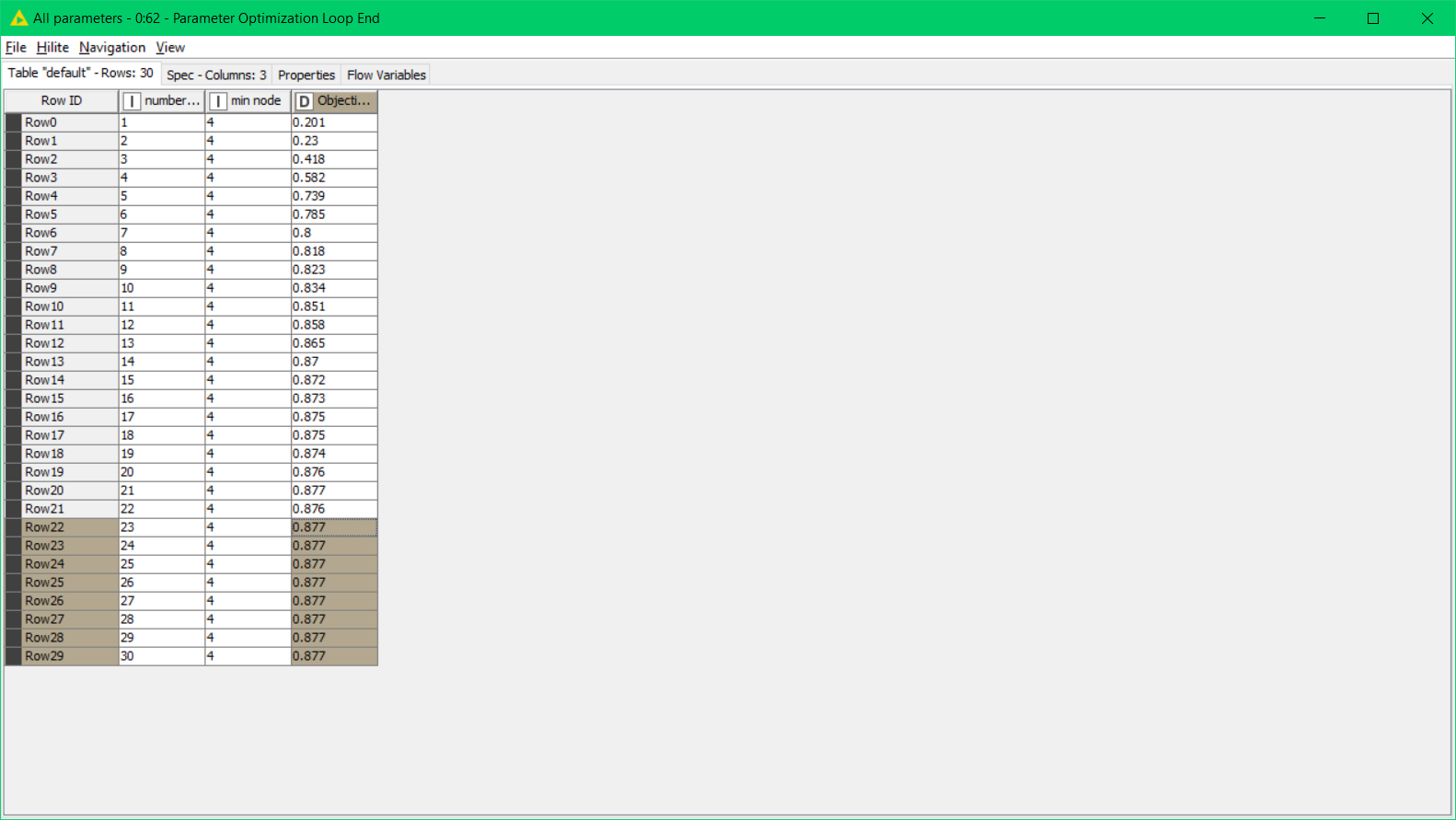
Retirar so avg precepitation e Roads M e Roads EM e Hora avg\_humidity



1 parametro

Selecionada

2 parametros



Mantendo o 1 parâmetro em 26

