

Auto Avaliação MVP – Thiago Coluccini

Através desse projeto, pude aprender muito sobre pyspark, SQL e Databricks, além de compreender a Medallion Architecture. Foi muito desafiador entender, documentar e desenvolver todo o projeto no tempo estimado, visto a complexidade da engenharia de dados.

Inicialmente, tinha como plano subir 2 bases de dados iguais, porém de períodos diferentes, para analisar divergências entre elas, como por exemplo:

- Os 10 melhores jogos avaliados em uma, são os mesmos que na outra?
- Quais jogos estavam com a melhor avaliação pela Steam em cada período?
- Quais jogos não existiam no período anterior e no período atual se sobressaíram nas avaliações?
- Quantos jogos foram lançados nessa diferença de tempo?

Porém, devido ao tempo de entrega do projeto, não foi possível realizar isso, devido a complexidade de modelagem de 2 tabelas FATOS, além de toda a criação de ETL e documentação de uma tabela adicional, tornando-se uma melhoria/implementação futura.

Além de que, ao longo do desenvolvimento, percebi que seria mais efetivo responder a esses questionamentos adicionais, com uma base mais antiga, de 5/10 anos atrás, visto que a diferença provavelmente seria maior, e teria mais conteúdo para ser analisado a partir disso.

Além de que, os resultados não foram exatamente como eu estava esperando, devido as avaliações da steam serem mais voláteis que as da crítica especializada (Metacritic) e apesar de tentar tratá-la de diversas formas conforme documentado no notebook e no PDF de evidências, para que não se sobressaísse, muitas vezes, tive a impressão de que os números dela ainda estavam “inflados”, talvez gerando uma divergência na análise, conforme foi mencionado nos documentos anteriores.

Sendo um possível ponto de atenção para melhoria futura, tentar entender e desenvolver um cálculo mais justo e preciso para as avaliações da Steam, definindo um peso para cada conceito.

No geral, consegui assegurar todos os questionamentos levantados, trazendo gráficos e explicações lógicas de como cheguei nos resultados determinados. Apesar de achar que poderia ter ido mais além nas análises, trazendo gráficos diferentes, análises mais aprofundadas e perguntas mais “fora da caixa”.

Apesar de tudo, acredito que fiz um ótimo trabalho, e consegui me desenvolver como pessoa, estudante e profissional, podendo dizer que domino o Databricks e gerando um excelente projeto para o portifólio.