JOB REQUIREMENTS

* coding and documentation best practices for MLOps
* Highly proficient in coding with python/pySpark
* Developing clean, well documented and tested code, production-level code and E2E automated pipelines
* data exploration, data quality assessment and cleansing of big volumes of data

TO REVISE

* Pyspark
* LGBM
* HYPEROPT
* STRUCTURE OF CODE AND GOOD PRACTICES
* MLFLOW
* MAKE AN ENVIRONMENT AND REQUIREMENTS TO USE DURING THE TEST
* MAKE THE PREPARATION TESTS ON HACKERRANK
* RECOMMENDATION SYSTEMS
* <https://christophm.github.io/interpretable-ml-book/pdp.html>
* GBM, CROSSVALIDATION, LIME, SHAP AND TIMESERIES
* <https://github.com/ModelOriented/DALEX>
* <https://dalex.drwhy.ai/python/api/fairness/index.html#dalex.fairness.GroupFairnessClassification.fairness_check>
* CHECK IMBALANCE PACKAGE
* CHECK HOW TO DEAL WITH OUTLIERS OR VERY SKEWED DISTRIBUTIONS
* STUDY SOME THEORY OF THE MOST ASKED DATA SCIENCE QUESTIONS

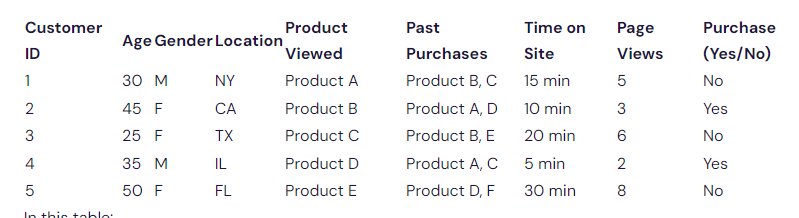
HINTS

* Gucci is using AI to help it understand consumer trends and preferences, as well as identify new product opportunities
* pattern recognition and the generation of timely and relevant counteractions
* AI in order to reduce waste by using it to inform systems about where and where not to sell products
* Watch this video <https://www.databricks.com/dataaisummit/session/mlops-gucci-zero-hero/>
* Get a free databricks subscription
* **Customer Profiling:** Gucci collected data from various touchpoints, including online interactions, purchase history, social media engagement, and in-store visits. This data was used to create detailed customer profiles, capturing preferences for products, styles, colors, and more.
* **Recommendation Engine:** The team implemented a recommendation engine that employed collaborative filtering and machine learning algorithms. This engine analyzed customer profiles and past behavior to suggest products that aligned with individual preferences. For instance, if a customer had shown interest in specific handbags, the engine could recommend complementary accessories.
* **Time series analysis**? Seasons?
* **Propensity to purchase - predictive behavioural model**

**TO REMEMBER IN THE INTERVIEW**

* Use well documented structured code
* Use hyperopt and mlflow

**FEATURES**

* If I need to build features for customers this might be some features to build (but more cleaned)****