**CHUẨN BỊ PHỎNG VẤN**

* **Các kiến thức chuyên môn đã ôn lại**
* Các loại chart, khi nào sử dụng các chart nào.
* Quy trình xử lý và làm sạch dữ liệu:

+ Thu thập dữ liệu

+ Khai phá dữ liệu (EDA)

+ Xác định và xử lý Missing value

+ Xử lý dữ liệu trùng lặp

+ Xử lý outlier

+ Chuẩn hoá dữ liệu

* Các cách phát hiện outlier và xử lý

+ Với dữ liệu numerical

+ Với dữ liệu categorical

+ Một số thuật toán phát hiện outlier: Kmean, HDBSCAN, Isolation forest, Local Outlier Factor, SVM one class

* Phân tích hồi quy
* Phân tích tương quan
* Phân tích phân cụm
* Các phương pháp kiểm định mô hình
* Kiểm định giả thuyết thống kê
* Mô hình phân tích đa chiều- OLAP

+ Quy trình xây dựng OLAP: Nghiệp vụ -> Xác định dim, fact -> Voi dim -> Data model logic, data model vật lý -> Phát triển các truy vấn và báo cáo

+ Phân tích với OLAP: Drill- down, Rollup, data slicing, data dicing

* Quy trình xây dựng một hệ thống
* Các sơ đồ UML: Sơ đồ phân cấp chức năng, ERD, sơ đồ usecase, sơ đồ lớp, sơ đồ tuần tự, sơ đồ hoạt động
* **Kết quả qua quá trình là thực tập sinh Data Analyst**
* Làm việc với mô hình Agile
* Học được cách làm việc hiệu quả và kinh hoạt trong môi trường đội nhóm
* Xây dựng một số KPI của TV360 ứng dụng vào màn chi tiết đối tượng trên hệ thống CCAI
* Học được cách xác định, xây dựng một KPI

+ Lựa chọn KPIs phải liên quan đến mục tiêu chính, mỗi KPI cần có mục tiêu và đo lường được

+ Các KPIs cần đo lường được sự tiến triển đối với mục tiêu của mình

+ Lựa chọn các KPIs mà cps thể kiểm soát và ảnh hưởng. Nếu không thể thay đổi hoặc ảnh hưởng đến một chỉ số -> nó có thể không phải là một KPI hiệu quả

* Xây dựng một số luồng tổng hợp dữ liệu cụ thể của màn chi tiết sự kiện và màn chi tiết đối tượng
* Cải thiện kỹ năng SQL, học được cách tối ưu hoá một luồng tổng hợp dữ liệu

+ Tối ưu hoá truy cập cơ sở dữ liệu: Chỉ trích xuất những dữ liệu cần thết

+ Lưu trữ các kết quả trước đó tránh việc tính toán lại các giá trị đã được tính toán trước đó

+ Giải phóng bộ nhớ khi không còn cần thiết

+ Chia nhỏ công việc thành các đơn vị nhỏ

+ Thiết lập một lịch trình cập nhật phù hợp với yêu cầu và tuần suất cập nhật của dữ liệu

* Hiểu về cách xử lý dữ liệu lớn sử dụng các công cụ như Hadoop, Spark
* Học được quy trình nghiệm thu các chức năng, thay đổi hàng tháng của hệ thống CCAI gồm:

+ Kế hoạch nghiệm thu: Ngày giờ, địa điểm, các biên bản và tài liệu liên quan

+ Đảm bảo đủ tài nguyên cho quá trình nghiệm thu bao gồm người kiểm thử, dữ liệu kiểm thử và môi trường kiểm thử

+ Kiểm tra các chức năng và thay đổi được triển khai trong hệ thống -> đảm bảo hoạt động đúng những yêu cầu không gây ra lỗi

+ Ghi lại và báo cáo mọi lỗi hoặc vấn đề phát sinh trong quá trình nghiệm thu

+ Sau khi tất cả các vấn đề được giải quyết -> triển khi trên môi trường thật

* Đầu ra cho quá trình này là cần đảm bảo hệ thống CCAI hoạt động, hiển thị đúng với từng chức năng như:

+ Màn Dashboard:

+ Màn Sự kiện:

+ Màn Insight Dashboard:

+ Màn Dự đoán: