Dàn bài project:

I. Data

1. Phân tích dataset

2. import data vào python notebook

3. Data issues:

- Fillna các value NA ở bảng general data = 0.0

- Merge các bảng manager\_survey và employee\_survey -> merge chung bảng general data

- drop các dòng có NA ở phần survey (4410 dòng -> 4300 dòng)

- melt hai bảng in\_time và out\_time -> merge thành một bảng lớn có hai cột in\_time và out\_time

4. Export data -> cleaned.csv, work\_time.csv

5. Tiếp tục clean data trong power bi:

- Tách bảng cleaned vào power bi thành 2 bảng, một bảng có các thông tin chung general\_data tên là cleaned và một bảng có các response của survey tên là survey.

- Tạo cột mới tên là AgeGroup để chia các nhân viên của công ty thành các nhóm tuổi: 18-24, 25-35, >36

- Drop column EmployeeCount.

- Trong bảng work\_time, đổi datatype của hai cột in\_time và out\_time thành time, thêm cột time\_at\_work = out\_time – in\_time

II. Phân tích data:

1. Overview: phân tích số lượng nhân viên của từng department, tỷ lệ nghỉ việc của năm qua, tỷ lệ nhóm tuổi, tỷ lệ giới tính của công ty.

2. Tạo measure Yes\_Attrition -> đếm số người có “Yes” ở cột Attrition. Sau đó tạo measure AttritionRate = tỷ lệ người nghỉ việc (AttritionRate = [Yes\_Attrition]/count(cleaned[EmployeeID]))

Lập các table để xét Attrition Rate theo từng nhóm (department, age group, gender, trình độ học vấn)

3. Lập các column chart để xem với những câu hỏi trong employee\_survey (EnvironmentSatisfaction, JobSatisfaction và WorkLifeBalance) được các nhân viên đã nghỉ việc trả lời như nào

4. Phân tích:

Work environment satisfaction thấp

* Avg = 2.47 -> chưa được tới trung bình (< 2.5)
* Xét những người có response < 3:
  + Tỷ lệ giới tính nam nữ ~ 60-40 -> tương đương tỷ lệ giới tính của cả công ty
    - Tuy nhiên, tỷ lệ này ở Department R&D là 55-45, Sales là 30-70
* Có thể có vấn đề liên quan đến giới tính ở R&D (do R&D là ngành nhiều nam giới hơn)
  + Years with current manager: 2.84 (của những người có nghỉ việc là 2.87), thấp hơn nhiều so với những người ở lại (trung bình là 4.38)
    - Nhóm 18-24: 0.82
    - Nhóm 25-35: 2.98
    - Nhóm 36 trở lên: 3.4
* Người trẻ không ở lâu, thời gian làm việc cùng manager hiện tại thấp hơn trung bình nhiều, có thể suy ra được là phong cách quản lý không hợp với những người trẻ (năm làm việc cùng của hai nhóm còn lại cao hơn trung bình của những nhóm này)
* Có thể có các yếu tố khác khiến cho hai nhóm còn lại nghỉ (thu nhập, các nhân viên khác, điều kiện làm việc)

Work Life balance:

* Avg = 2.66 -> above average
* Xét những người có response < 3 (Low/Average)
  + Tỷ lệ overtime (làm trên 9 tiếng) của những nhân viên phải làm overtime (overtime\_rate >= 0.1) là 0.86
  + Avg Distance from home ~ 9.4 -> xa
* Cơ sở hạ tầng giao thông của Ấn Độ không tốt (tắc đường, nhiều ô tô), kèm với việc phải làm overtime làm cho người nhân viên không có đời sống cân bằng
* https://www.tomtom.com/traffic-index/india-country-traffic/
  + Xét tiếp nhóm tuổi, tình trạng hôn nhân: nửa 25-35, 12% 18-24 tuổi (nhóm tuổi), gần nửa là single, 1/3 đã kết hôn. Phần lớn là phải đi công tác (rarely), 24% là phải đi thường xuyên.
  + Xét trong nhóm đã kết hôn, phân bố tỷ lệ cũng tương tự.
* Việc phải đi công tác ảnh hưởng nhiều tới những nhân viên đã kết hôn, do nó giảm thời gian ở bên gia đình.
* Solution: Cho thêm lựa chọn làm hybrid/remote cho những người ở xa nhà quá, giảm tần suất đi công tác của mỗi nhân viên, giảm overtime

Job satisfaction thấp -> để ý các factor liên quan (yrs since last promo, salary compared to avg)

* Avg = 2.47 (<2.5)
* <https://www.statista.com/statistics/654471/salary-employee-average-monthly-income-india/> (distribution thu nhập hàng tháng ở Ấn Độ năm 2015) -> chủ yếu là ở mức 20001-50000 rupee, thu nhập trên 50000 nhiều hơn so với mặt bằng chung
* Salary Hike của cty cao hơn thực tế (tại năm 2015, salary hike là 10.4%)
* Với những response < 3:
  + phần lớn là những người job level thấp nhưng ở công ty lâu (trên 25 tuổi nhưng job level 1 và 2) -> chán nản do không thăng tiến được làm cho jobsatisfaction thấp

Table

Description automatically generated

* + lương của level 5 thấp -> lương không tương xứng với job level

Table

Description automatically generated

* solution: mở ra các cơ hội thăng tiến nhiều hơn cho nhân viên, trả lương tương xứng với level