**Bài 1:** Dựa vào dữ liệu kèm theo thực hiện 1 số yêu cầu sau:

1. Tính tổng tiền hóa đơn và tiền tip trung bình theo từng ngày.
2. Tính tỷ lệ tiền tip so với tổng hóa đơn, sau đó tìm khách hàng nào đã tip nhiều nhất (theo tỷ lệ).
3. So sánh số lượng khách hàng nam và nữ, xem giới tính nào tip nhiều hơn trung bình.
4. Tạo một cột mới là "Tip Percentage" để biểu diễn tỷ lệ tip so với tổng hóa đơn, sau đó lọc ra những bữa ăn có "Tip Percentage" lớn hơn 20%.
5. Tính trung bình số người tham gia bữa ăn và tổng tiền tip cho các bữa ăn diễn ra vào cuối tuần (Sat, Sun), khách hàng không hút thuốc, và là bữa tối (Dinner).

**Bài 2:** Dựa vào dữ liệu kèm theo thực hiện 1 số yêu cầu sau:

1. Đọc dữ liệu và xem xét các cột có giá trị thiếu (NaN), tính toán tỷ lệ phần trăm dữ liệu bị thiếu cho mỗi cột.
2. Xử lý các giá trị bị thiếu trong dữ liệu. Bạn có thể chọn cách điền giá trị trung bình của cột hoặc loại bỏ hàng có dữ liệu bị thiếu.
3. Tạo biểu đồ cột (bar chart) để hiển thị tỷ lệ phần trăm dữ liệu bị thiếu của từng cột.
4. Tìm những khoảng thời gian mà chỉ số PM2.5 vượt quá ngưỡng 100 (chỉ số chất lượng không khí rất xấu).
5. Tạo biểu đồ đường (line chart) hiển thị sự thay đổi của PM2.5 theo thời gian, và làm nổi bật những thời điểm vượt quá ngưỡng 100.
6. Xác định khoảng thời gian dài nhất mà chất lượng không khí ở mức xấu, tức là PM2.5 liên tục trên 100.
7. Tạo một cột mới là "ngày" từ cột Datetime để phân tích theo từng ngày.
8. Tính chỉ số trung bình của các chất ô nhiễm chính (PM2.5, PM10, NO2) theo từng ngày.
9. Tạo biểu đồ đường (line chart) để hiển thị xu hướng biến động của PM2.5 theo ngày trong toàn bộ khoảng thời gian.
10. Tạo một cột mới từ Datetime đại diện cho "ngày trong tuần" (Thứ 2, Thứ 3, v.v.).
11. Tính trung bình của các chất ô nhiễm theo từng ngày trong tuần.
12. Xác định ngày nào trong tuần có mức độ ô nhiễm cao nhất.

**Bài 3:** Dựa vào dữ liệu kèm theo thực hiện 1 số yêu cầu sau:

1. Tính lương trung bình của tất cả nhân viên.Tính lương trung bình theo giới tính (Gender).
2. Vẽ đồ thị: Sử dụng biểu đồ cột (bar chart) để hiển thị lương trung bình của nhân viên theo giới tính.
3. Tạo một cột mới trong DataFrame tính toán tổng tiền thưởng cho mỗi nhân viên (tính bằng lương nhân với phần trăm thưởng).
4. Tìm nhân viên có tổng tiền thưởng cao nhất và thấp nhất.
5. Vẽ đồ thị: Sử dụng biểu đồ cột để hiển thị 10 nhân viên có tổng tiền thưởng cao nhất.
6. Tạo cột mới tính số năm làm việc của mỗi nhân viên (dựa vào Start Date và ngày hiện tại).
7. Tính số năm làm việc trung bình của nhân viên theo phòng ban (Team).
8. Vẽ đồ thị để hiển thị sự phân phối số năm làm việc của nhân viên theo phòng ban.
9. Tính số lượng nhân viên trong mỗi phòng ban (Team). Tính tỷ lệ phần trăm số lượng nhân viên trong mỗi phòng ban.
10. Vẽ đồ thị: sử dụng biểu đồ tròn (pie chart) để hiển thị tỷ lệ phần trăm số lượng nhân viên của từng phòng ban.

**Bài 4:** Dựa vào dữ liệu kèm theo thực hiện 1 số yêu cầu sau:

1. Tính tỉ lệ phần trăm những người sống sót và không sống sót.Vẽ đồ thị: Sử dụng biểu đồ cột (bar chart) hoặc biểu đồ tròn (pie chart) để hiển thị tỉ lệ sống sót (Survived).
2. Tính tỉ lệ sống sót của hành khách theo giới tính (Sex).Vẽ đồ thị: Sử dụng biểu đồ cột nhóm (grouped bar chart) để so sánh tỉ lệ sống sót giữa nam và nữ.
3. Tính tỉ lệ sống sót theo hạng vé (Pclass). Vẽ đồ thị: Sử dụng biểu đồ cột để hiển thị tỉ lệ sống sót của hành khách theo từng hạng vé (1, 2, 3).
4. Tính độ tuổi trung bình, nhỏ nhất và lớn nhất của các hành khách. Vẽ đồ thị: Vẽ biểu đồ histogram để hiển thị phân bố độ tuổi của hành khách.
5. Phân chia hành khách thành các nhóm tuổi (trẻ em: 0-12, thanh thiếu niên: 13-19, người lớn: 20-60, người già: 60+). Tính tỉ lệ sống sót của mỗi nhóm tuổi.
6. Vẽ đồ thị: Sử dụng biểu đồ cột để hiển thị tỉ lệ sống sót của từng nhóm tuổi.
7. Tạo cột mới để tính tổng số người thân đi cùng cho mỗi hành khách (SibSp + Parch).
8. Tính tỉ lệ sống sót của hành khách đi một mình và đi cùng gia đình.
9. Vẽ đồ thị: Vẽ biểu đồ cột để so sánh tỉ lệ sống sót giữa hành khách đi một mình và đi cùng gia đình.