



Data Analytics Bootcamp

Bahrain – General Assembly

The Hidden Health Effects of Smoking

Prepared By: Kawthar Isa

DAB-9

Table of Contents

1. Introduction.....	3
2. Problem Statement	4
3. Objectives	4
4. Target Audience	4
5. Dataset(s).....	5
6. Data Handling.....	7
7. Analysis and Findings	8
8. Recommendations.....	8
9. Limitations and Assumptions.....	9
10. References.....	9

1. Introduction

Smoking remains one of the most preventable causes of disease and premature death worldwide. It introduces thousands of harmful chemicals into the body, many of which are toxic and carcinogenic. The most well-documented and immediate target of smoking is the **respiratory system**, particularly the lungs. With every inhalation of cigarette smoke, the delicate tissues of the lungs are exposed to tar, carbon monoxide, and other toxins that damage airways, reduce lung capacity, and destroy the tiny air sacs (alveoli) essential for oxygen exchange.

Chronic exposure leads to **serious lung diseases**, such as:

- **Chronic Obstructive Pulmonary Disease (COPD)** – including chronic bronchitis and emphysema.
- **Lung cancer** – the leading cause of cancer death globally, strongly linked to tobacco use.
- **Asthma exacerbation** and increased respiratory infections.

However, the damage from smoking is not limited to the lungs. Smoking affects nearly every organ in the body and is associated with a wide range of **systemic health conditions**, including:

- **Cardiovascular diseases** such as heart attack and stroke.
- **Metabolic disorders** like type 2 diabetes and metabolic syndrome.
- **Liver and kidney damage**, including increased risks of cirrhosis and chronic kidney disease.
- **Sensory impairments**, particularly hearing loss and vision decline.

The widespread impact of smoking underscores the importance of comprehensive prevention and cessation programs, not only to protect the lungs but to preserve overall health and quality of life.

2. Problem Statement

Smoking is a leading cause of preventable diseases and premature death globally. While its detrimental effects on the lungs are well-known, the broader medical implications of smoking across various body systems are often overlooked. This study aims to explore the extended health risks associated with smoking by analyzing its relationship with multiple chronic diseases using real-world population data.

3. Objectives

- Investigate the impact of smoking on the incidence and severity of multiple chronic diseases.
- Identify disease-specific risk levels among smokers versus non-smokers.
- Present visual insights into the distribution of health conditions by smoking status.

4. Target Audience

- Medical researchers
- Public health policymakers
- Clinical professionals
- Anti-smoking advocacy groups

5. Dataset(s)

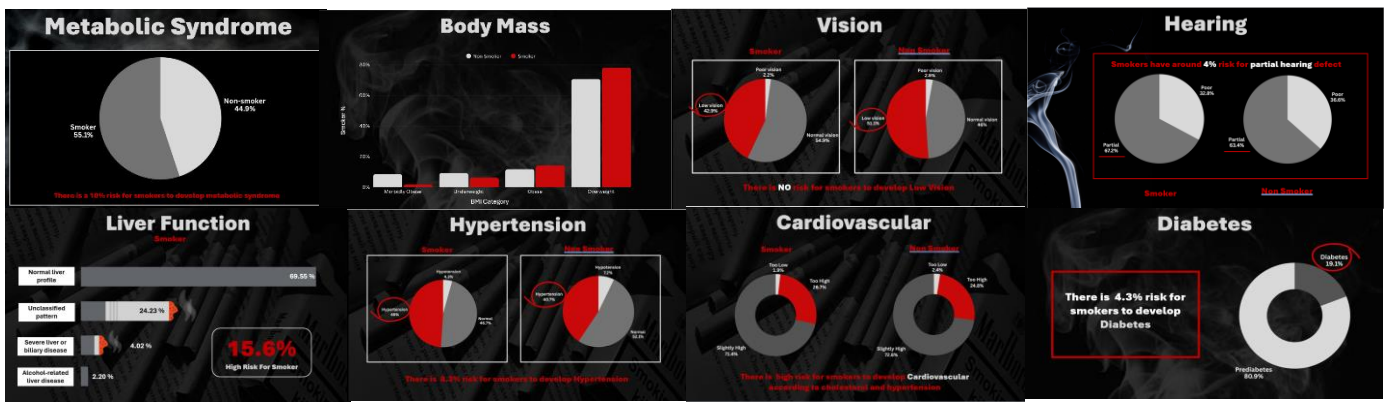
This dataset from Kaggle contains medical examination records used to analyze the effects of smoking on various health indicators. It includes measurements such as blood pressure, cholesterol, blood sugar, liver enzymes, vision, hearing, BMI, and more. The data was collected as part of regular health checkups in South Korea, targeting a large sample of adults. Each record is labeled with smoking status, allowing for comparative analysis between smokers and non-smokers. This dataset supports research into how smoking impacts different body systems beyond just the lungs.

Column Name	Description
ID	Unique identifier for each record
gender	Gender of the individual (e.g., Male, Female)
age	Age of the individual (in years)
height(cm)	Height in centimeters
weight(kg)	Weight in kilograms
waist(cm)	Waist circumference in centimeters
eyesight(left)	Left eye vision measurement
eyesight(right)	Right eye vision measurement
hearing(left)	Hearing test result for the left ear (0 = Normal, 1 = Abnormal)
hearing(right)	Hearing test result for the right ear
systolic	Systolic blood pressure (mmHg)
relaxation	Diastolic blood pressure (mmHg)
fasting blood sugar	Blood glucose level after fasting (mg/dL)
Cholesterol	Total cholesterol (mg/dL)
triglyceride	Triglycerides level (mg/dL)
HDL	High-density lipoprotein (good cholesterol) (mg/dL)
LDL	Low-density lipoprotein (bad cholesterol) (mg/dL)
hemoglobin	Hemoglobin concentration in blood (g/dL)
Urine protein	Urinary protein levels (0 = Negative, 1+ to 4+ = Positive range)
serum creatinine	Serum creatinine level (mg/dL), indicates kidney function
AST	Aspartate transaminase level (U/L), liver enzyme
ALT	Alanine transaminase level (U/L), liver enzyme
Gtp	Gamma-glutamyl transferase level (U/L), liver enzyme
oral	Oral health check result (text format)
dental caries	Number of decayed teeth (caries count)
tartar	Presence of tartar (Yes/No or similar)
smoking	Smoking status (missing in data, may be derived)
AST_status	Classification of AST (e.g., Normal, High)
ALT_status	Classification of ALT

GGT_status	Classification of GTP (GGT enzyme)
Liver_status	Liver health assessment summary
creatinine_status	Classification of serum creatinine levels
urine_protein_status	Categorized status of urine protein
kidney_status	Kidney health assessment summary
BMI	Body Mass Index (kg/m ²)
BMI_category	Classification of BMI (e.g., Underweight, Normal, Overweight, Obese)
vision_status	Summary status of eyesight (e.g., Normal, Impaired)
Hearing Status	Summary status of hearing (e.g., Normal, Defect)
blood_pressure_status	Classification of blood pressure (e.g., Normal, High)
FBS_status	Fasting Blood Sugar status (e.g., Normal, Prediabetic, Diabetic)
Cholesterol_status	Classification of cholesterol levels
Hemoglobin_status	Hemoglobin level status (e.g., Normal, Low, High)
waist_risk	Indicates whether waist circumference poses a metabolic risk
LDL_status	Classification of LDL levels
HDL_status	Classification of HDL levels
triglyceride_status	Classification of triglyceride levels
features_list	Combined or derived health feature categories
metabolic_syndrome	Final label indicating presence of metabolic syndrome (Yes/No)

6. Data Handling

Column Name	Data Handling Description
AST_status	Categorized AST values into: 'Low level', 'Normal', 'High level', or 'Unknown'.
ALT_status	Categorized ALT values similarly into severity levels.
GGT_status	Classified GGT (Gtp) into: 'Normal', 'High', or 'Very High / Abnormal'.
Liver_status	Combined AST, ALT, and GGT status to generate a liver health interpretation.
creatinine_status	Classified serum creatinine based on gender into 'Low', 'Normal', or 'High' levels.
urine_protein_status	Classified Urine protein levels into severity ranges or marked as 'unknown'.
kidney_status	Derived overall kidney function by combining creatinine and urine protein statuses.
BMI	Calculated BMI from height(cm) and weight(kg).
BMI_category	Categorized BMI into: Underweight, Normal, Overweight, Obese, or Morbidly Obese.
vision_status	Evaluated vision quality based on eyesight(left) and eyesight(right).
Hearing Status	Assessed hearing status using hearing(left) and hearing(right) columns.
blood_pressure_status	Categorized blood pressure using systolic and relaxation (diastolic) values.
FBS_status	Classified fasting blood sugar into: Normal, Prediabetes, or Diabetes.



7. Analysis and Findings

Metabolic Syndrome

- 10% of smokers are at risk of developing metabolic syndrome.

Vision Impairment

- There is no risk for smokers to develop low vision.

Hearing Defect

- Smokers show a 4% risk of partial hearing loss.

Liver Function Impairment

- 15.6% of smokers are at high risk for abnormal liver function.

Hypertension

- 8.3% of smokers are at risk of developing hypertension.

Diabetes

- 4.3% of smokers are at risk of developing diabetes.

Cardiovascular Disease

- Smokers show a high overall risk for cardiovascular disease due to elevated cholesterol and hypertension levels.

8. Recommendations

- 1- STOP smoking

9. Limitations and Assumptions

- **Cross-sectional Data:** The study uses cross-sectional data, which limits causal inference.
- **Uncontrolled Confounders:** Other variables such as diet, alcohol use, and genetics were not analyzed.
- **Self-Reporting Bias:** Smoking status may be underreported by participants.
- **Scope Limitation:** Focused on specific diseases and did not include cancer or respiratory diseases.

10. References

- World Health Organization: Health effects of smoking
- CDC: Smoking-related health conditions
- Peer-reviewed journals on metabolic syndrome, hypertension, and cardiovascular disease in smokers