Vision Care Insurance :
Key Factors to design policies

Team : Section 001(Group 2)

Ikram Patel

Rituraj Borah
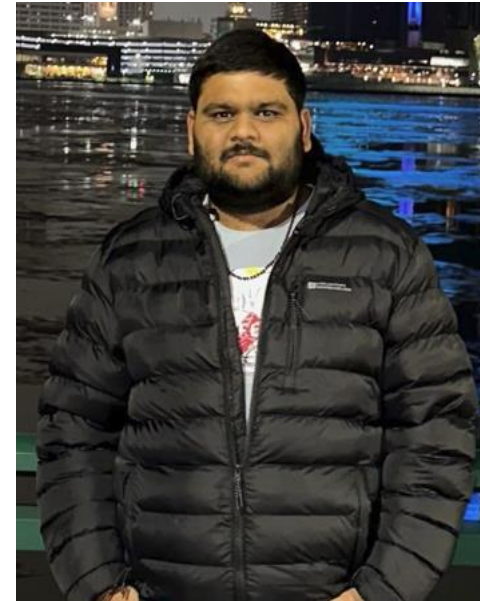
Tirthak Bhingaradiya

LinkedIn

Github

LinkedIn

Github

LinkedIn

Github

OBJECTIVE

1. Background Motivation
2. Problem Statement
3. Project Proposal
4. Analysis Questions
5. Dataset Description
6. Data Cleaning & Transformation
7. Analysis Results

BACKGROUND MOTIVATION

➢ Vision is critical for daily tasks, but due to screen time, diabetes, and aging, vision problems are common nowadays.

➢ The World Health Organization estimates that approximately 2.2 billion people globally have some form of vision impairment, with the majority being easily preventable or treatable *("Vision impairment and blindness," 2021).*

➢ As per future projections made by most recent census data, the number of people with visual impairment or blindness in the United States is expected to double to more than 8 million by 2050.*("Visual impairment, blindness cases in U.S. expected to double by 2050," n.d.)*

BACKGROUND MOTIVATION…

➢ Health insurance providers offer coverage plans to assist customers in dealing with the financial burden associated with critical illnesses

➢ It offers coverage for critical illnesses, including vision care. Vision insurance typically covers routine eye exams, glasses, contacts, and other related services.

➢ Since there are risk factors such as Race/ethnicity, age, high blood pressure, heart disease, smoking, and physical inactivity which can increase the risk of vision loss.

➢ Insurance companies aim to provide the most superior eye-care insurance plans in the market to motivate their customers to try out new or enhanced products.

➢ Nevertheless, the development of these new offerings demands comprehensive research and analysis of the previous health-related challenges experienced by their intended customers, considering various factors.

PROBLEM STATEMENT

Risk factors

Better insights

Pre-existing disease

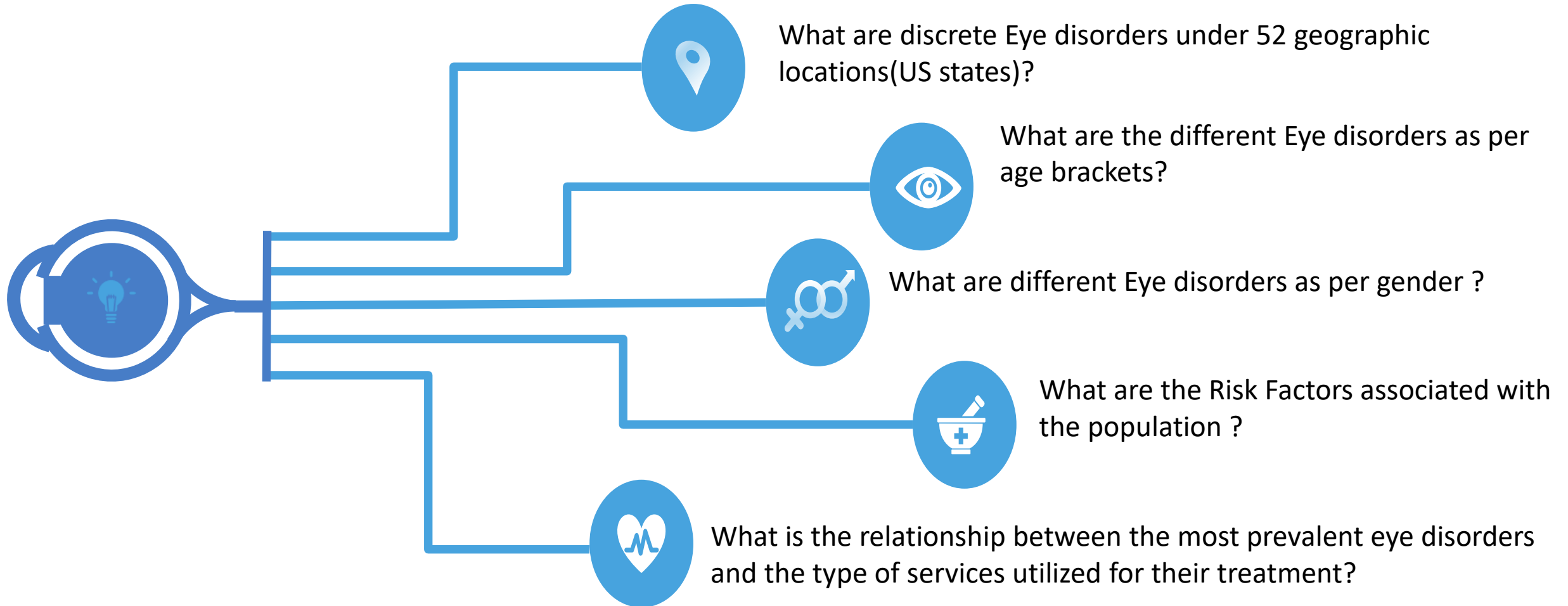Current trends

Effective Insurance plans
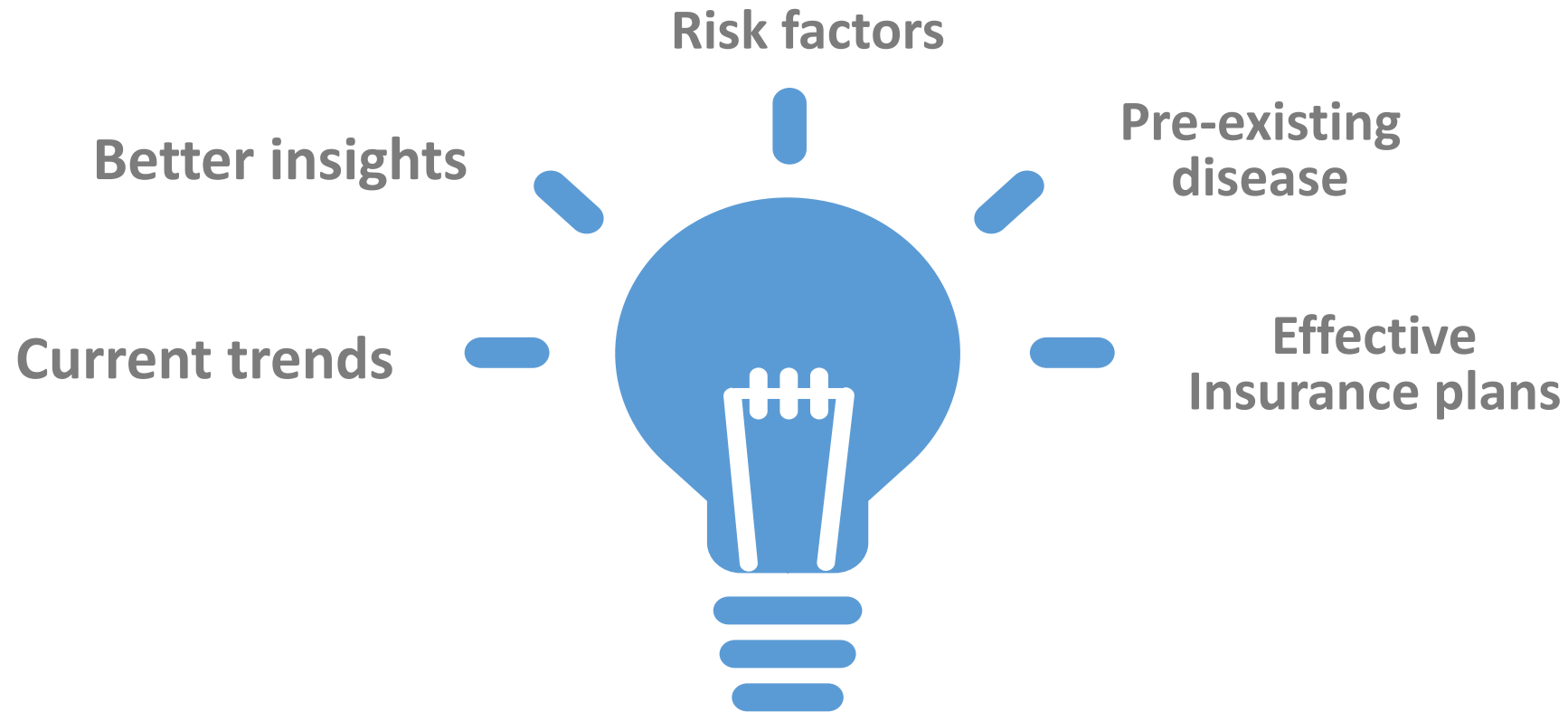
In order to create vision insurance policies that are more tailored to the needs of their consumers and generate revenue, insurance companies in North America, particularly the US, require a more comprehensive understanding of current trends and risk factors in eye care specific to different age groups and pre-existing medical conditions of the population.

PROJECT PROPOSAL

➢ Our team will present an analysis of patterns and fluctuations observed in the data obtained from the Centre for Disease Control and Prevention.

➢ The study will support the claims made regarding different geographical areas, age groups, and types of eye ailments.

➢ This data comprises a summarized table of prevalence rates for various indicators related to vision and eye health, extracted from the year 2016.

➢ Our research will provide valuable insights to insurance companies, assisting them in designing and recommending suitable products to customers.

➢ It will also aid in making financial decisions such as insurance amount and premiums based on various factors, enabling them to make a profit.

ANALYSIS QUESTIONS

What are discrete Eye disorders under 52 geographic locations(US states)?

What are the different Eye disorders as per age brackets?

What are different Eye disorders as per gender ?

What are the Risk Factors associated with the population ?

What is the relationship between the most prevalent eye disorders and the type of services utilized for their treatment?

DATASET DESCRIPTION

➢ This dataset contains a summary table of prevalence rates for vision and eye health data indicators from the year 2016.

➢ Data source: Commercial Claims and Encounters Data (CCAE) is produced by Truven Health Analytics, a division of IBM Watson Health.

➢ *The dataset is collected by doing a survey for an equal number of people across each Gender and Age category for various types of diseases across the states in the US*. Thus, the data can be considered unbiased and a uniform representation of the population.

➢ The data contain a convenience sample of insurance claims, and the data is arranged by all available combinations of age group, gender, and state.

➢ As per our initial assessment, there are enough variables/columns to perform a detailed analysis and share insight from the dataset.

➢ The data preparation will undergo a few minor transformations before projecting it for analysis.

DATASET DETAILS

Column Name	Description	Type
YearStart	Starting year for year range	Number
YearEnd	Ending year for year range. Same as starting year if single year used in evaluation	Number
LocationAbbr	Location (National or state) abbreviation	Text
LocationDesc	Location (National or state) full name	Text
DataSource	Abbreviation of Data Source	Text
Topic	Topic description	Text
Category	Category description	Text
Question	Question description (e.g., Percentage of adults with diabetic retinopathy)	Text
Response	Optional column to hold the response value that was evaluated	Text
Age	Stratification value for age group (e.g., 0-17 years)	Text
Gender	Stratification value for gender (e.g., Total, Male, Female, or Unknown)	Text
RaceEthnicity	Stratification value for race (e.g., All races, Asian, Black, non-hispanic, Hispanic, any race, North American Native, White, non-hispanic, or Other)	Text
RiskFactor	Stratification value for major risk factor (e.g., All participants, Diabetes)	Text
RiskFactorResponse	Column holding the response for the risk factor that was evaluated (e.g., All participants, Yes, No, or Unknown)	Text
Data_Value_Unit	The unit, such as "%" for percent	Text
Data_Value_Type	The data value type, such as age-adjusted prevalence or crude prevalence	Text
Data_Value	A numeric data value greater than or equal to 0, or no value when footnote symbol and text are present	Number

DATASET DETAILS

Column Name	Description	Type
Data_Value_Footnote_Symbol	Footnote symbol	Text
Data_Value_Footnote	Footnote text	Text
Low_Confidence_Limit	95% confidence interval lower bound	Number
High_Confidence_Limit	95% confidence interval higher bound	Number
Numerator	The actual number of patients with the condition (n)	Number
Sample_Size	Sample size used to calculate the data value	Number
LocationId	Lookup identifier value for the location	Text
TopicId	Lookup identifier for the Topic	Text
CategoryId	Lookup identifier for the Category	Text
QuestionId	Lookup identifier for the Question	Text
ResponseId	Lookup identifier for the Response	Text
DataValueTypeId	Lookup identifier for the data value type	Text
AgeId	Lookup identifier for the Age stratification	Text
GenderId	Lookup identifier for the Gender stratification	Text
RaceEthnicityId	Lookup identifier for the Race/Ethnicity stratification	Text
RiskFactorId	Lookup identifier for the Major Risk Factor	Text
RiskFactorResponseId	Lookup identifier for the Major Risk Factor Response	Text
Geolocation	Latitude & Longitude to be provided for formatting GeoLocation in the format (latitude, longitude)	Point

VARIABLES IN DATASET

LocationAbbr
LocationDesc
Topic
Category
Response
Age
Gender
RiskFactor

Categorical Variables

Additional Variables
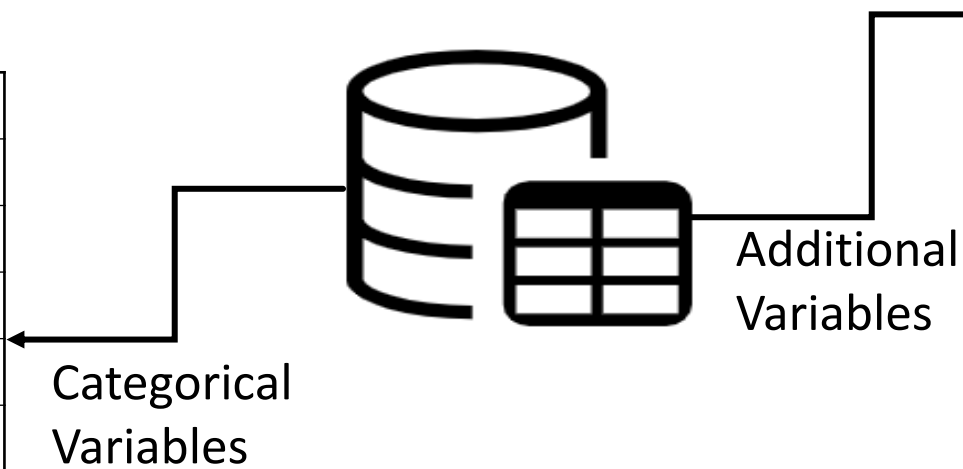
YearStart
YearEnd
RaceEthnicity
RiskFactorResponse
Data_Value_Type
Low_Confidence_Limit
High_Confidence_Limit
Numerator
Sample_Size
LocationId
TopicId
CategoryId
QuestionId
ResponseId
DataValueTypeId
AgeId
GenderId
RaceEthnicityId
RiskFactorId
RiskFactorResponseId
Geolocation

EXPLORATORY DATA ANALYSIS

➢ Locating a trustworthy source for the dataset.

➢ Examining the data frame information to comprehend the number of rows and columns, data types, and non-null values.

➢ Removing irrelevant data values, such as filtering out two incorrect location values.

➢ Saving the dataset on Google Drive to commence working on Google colaboratory.

➢ Managing failures and errors, if any occur.

DATA CLEANING & TRANSFORMATION

➢ We utilized Python along with the Pandas library to import and merge data from two source files, and to facilitate the exploration of the data.

➢ We have cleaned the data in Location column by removing irrelevant or inaccurate values.

➢ Values in Category column has been transformed to make them more concise and visually appealing for better data visualization.

➢ As there were 52 discrete locations, it was challenging to showcase them in plots. Therefore, we categorized the locations into four different regions to improve data visualization.

ANALYSIS 1 : Distribution of Eye disorders across Locations
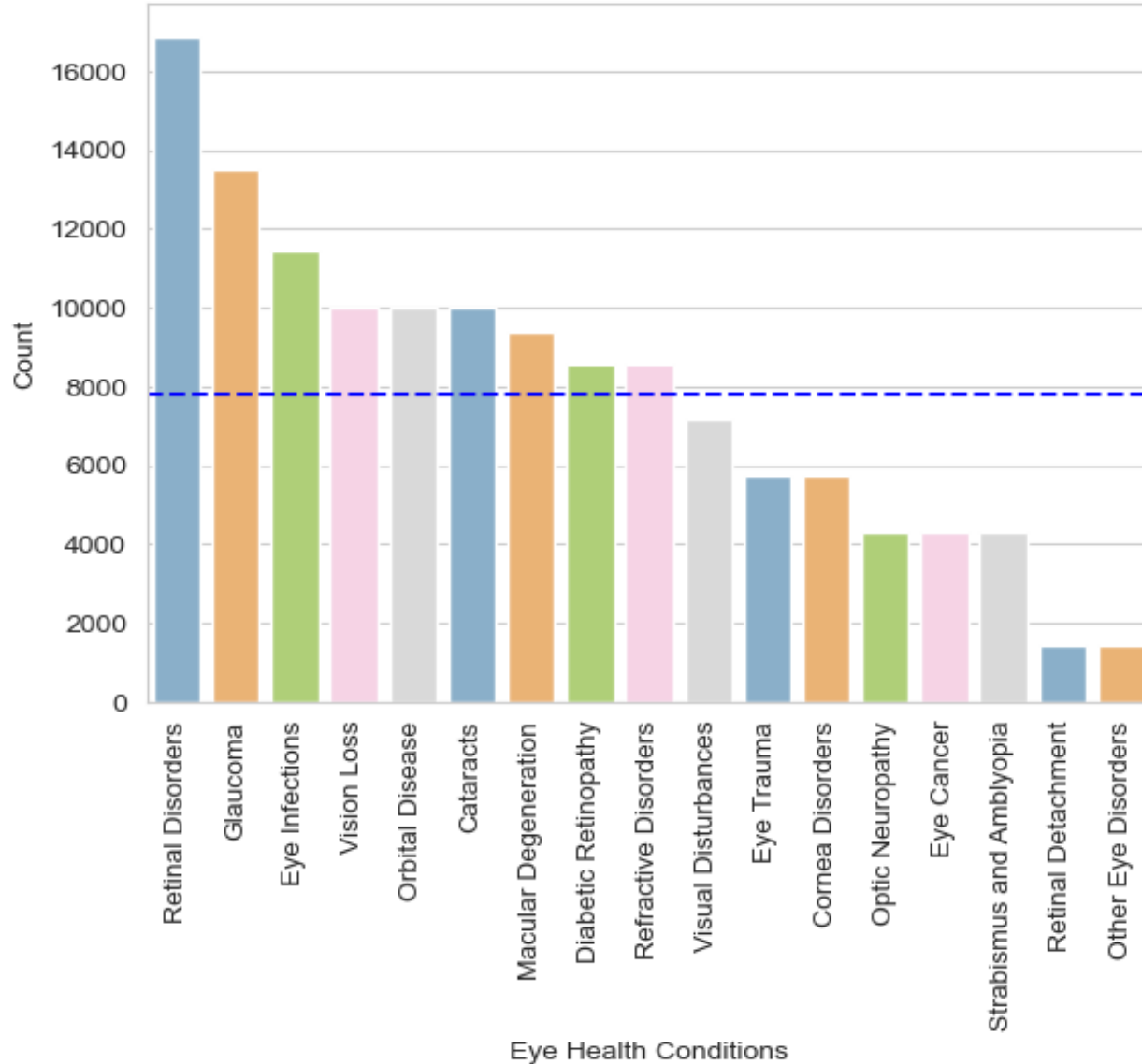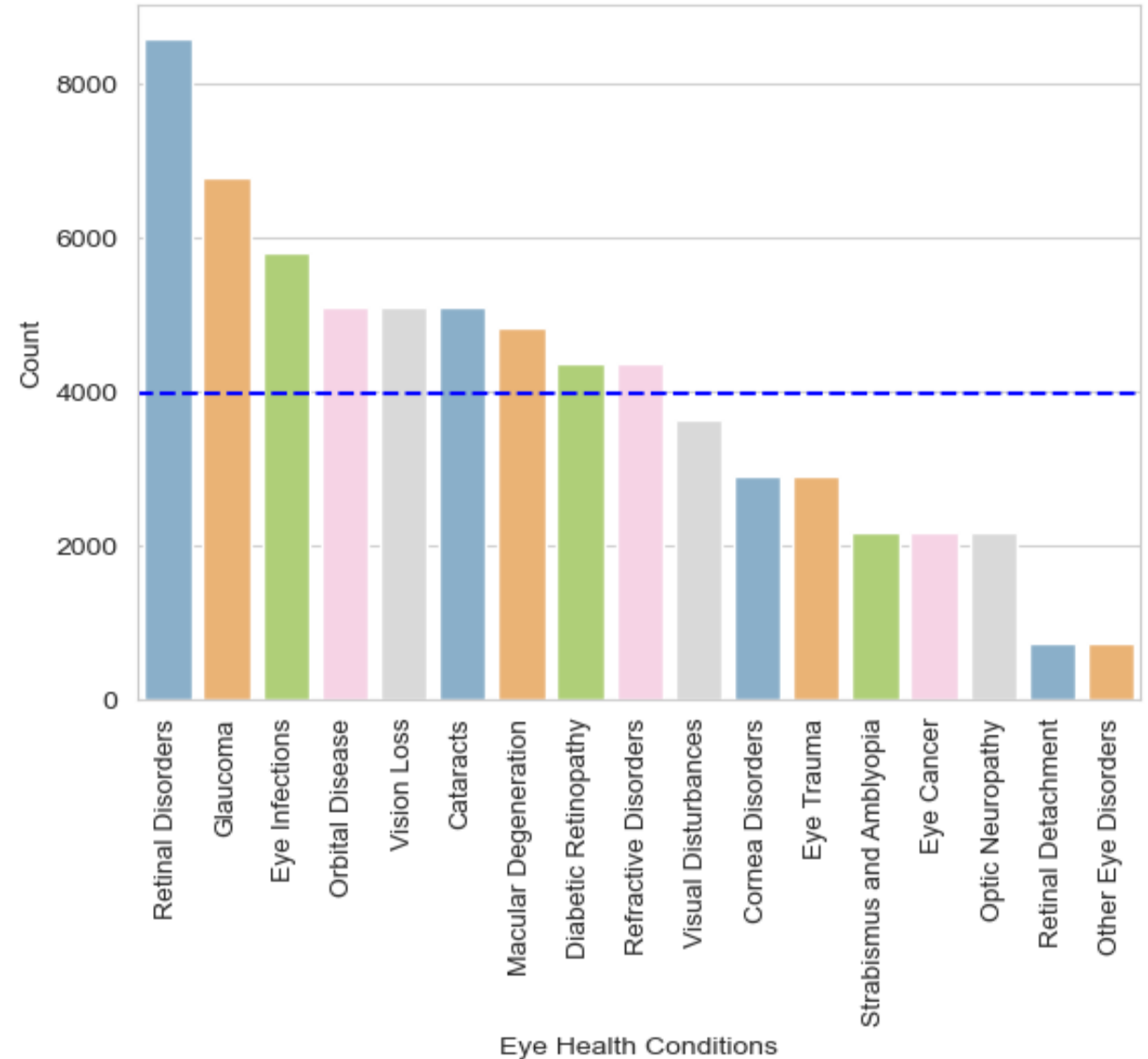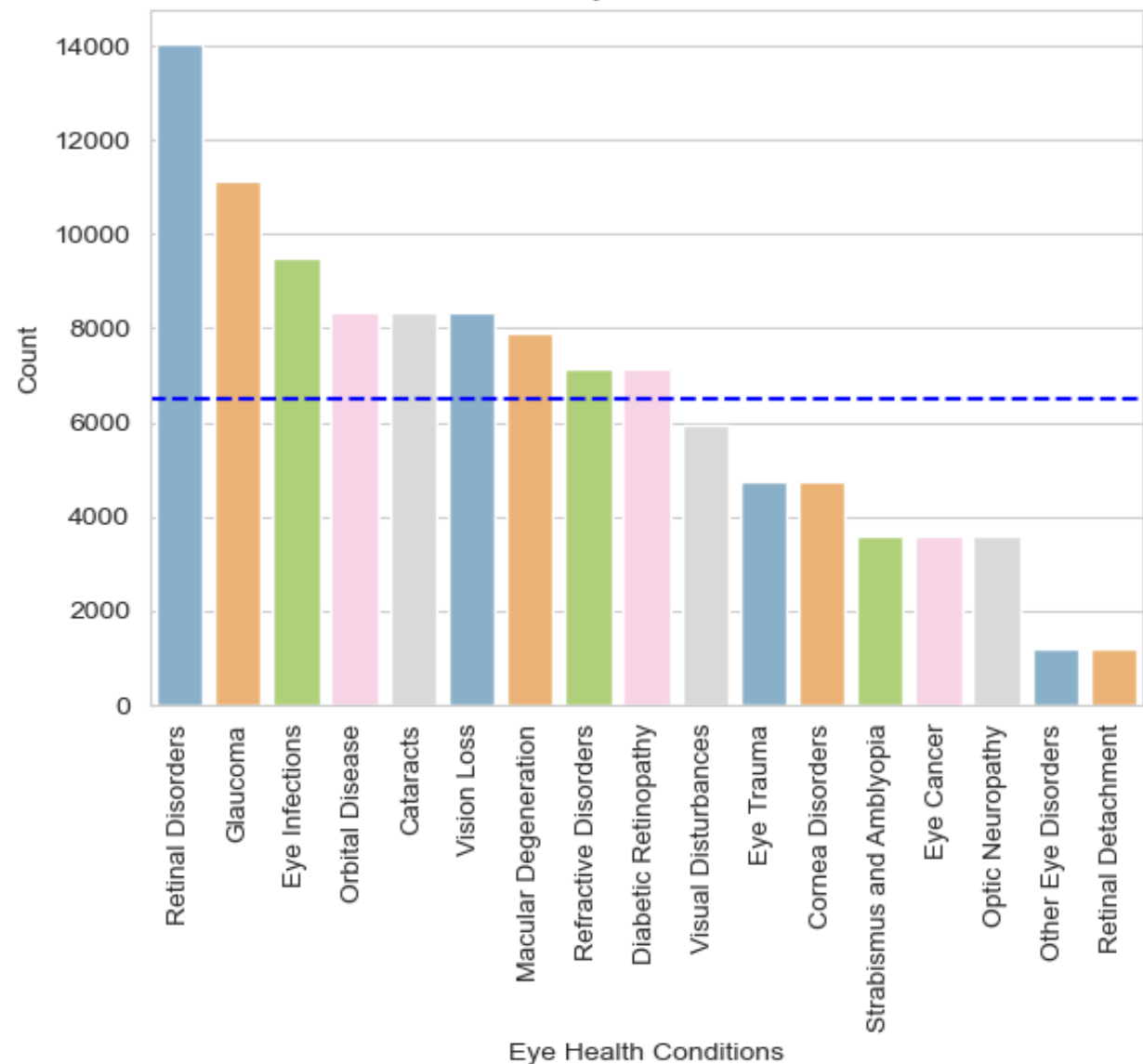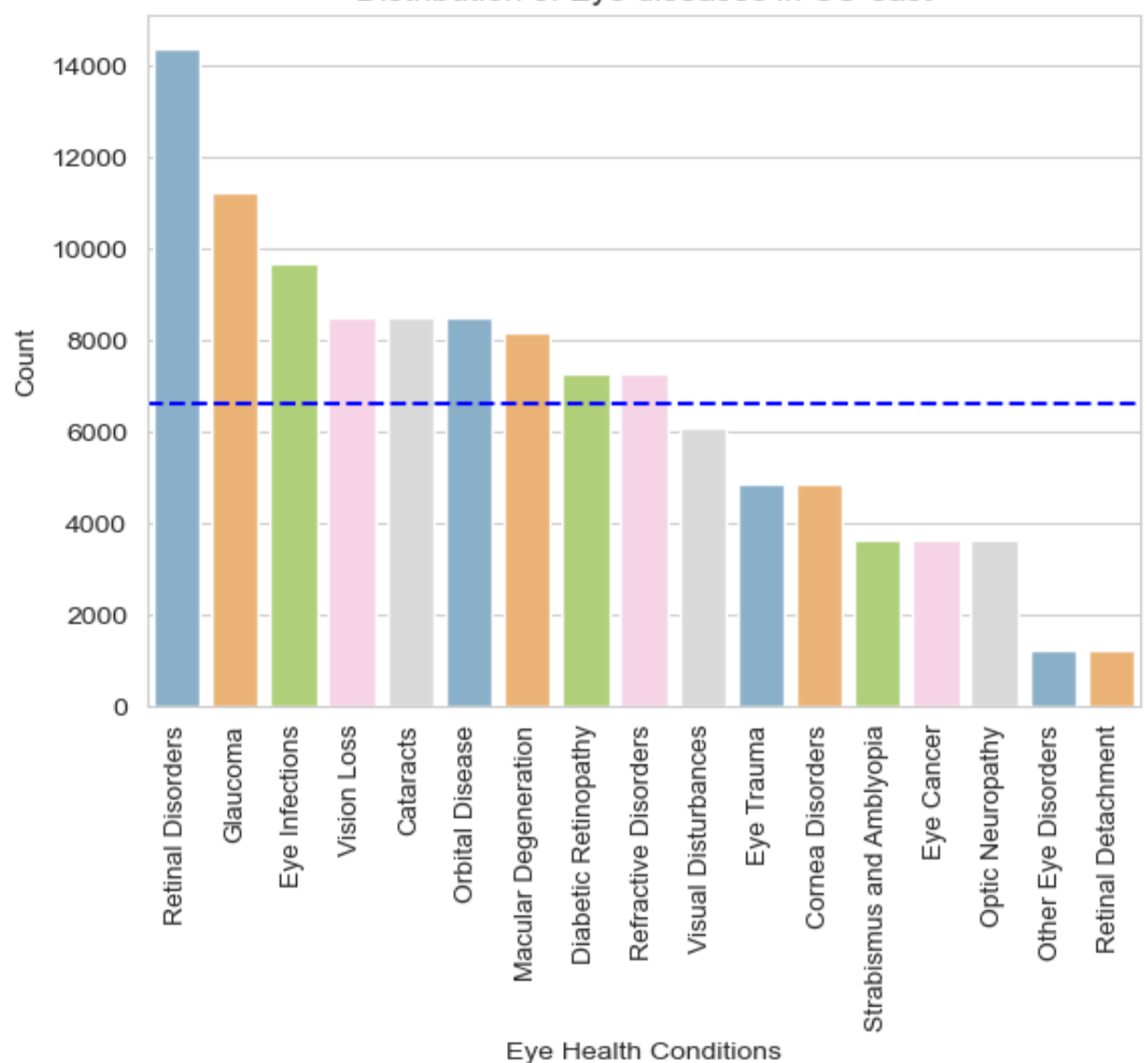
# ANALYSIS 1 : Distribution of Eye disorders across Locations...

➢ The graph provides a clear overview of the common vision-related issues across various regions in the US.

➢ **Retinal Disorders** and **Glaucoma** are the most common vision-related issues in all of them.

➢ Several other eye diseases such as *Eye infections*, *Vision Loss*, *Orbital Diseases*, *Cataracts*, *Macular Degeneration*, *Refractive Disorders*, and *Diabetic Retinopathy* are prevalent among the population and affect more than the average population in all regions.

➢ The population in **US-north** has the lowest prevalence of vision-related issues compared to other regions.

ANALYSIS 2 : Distribution of Eye disorders by Age Groups

Distribution of Eye Health Conditions by Age Group
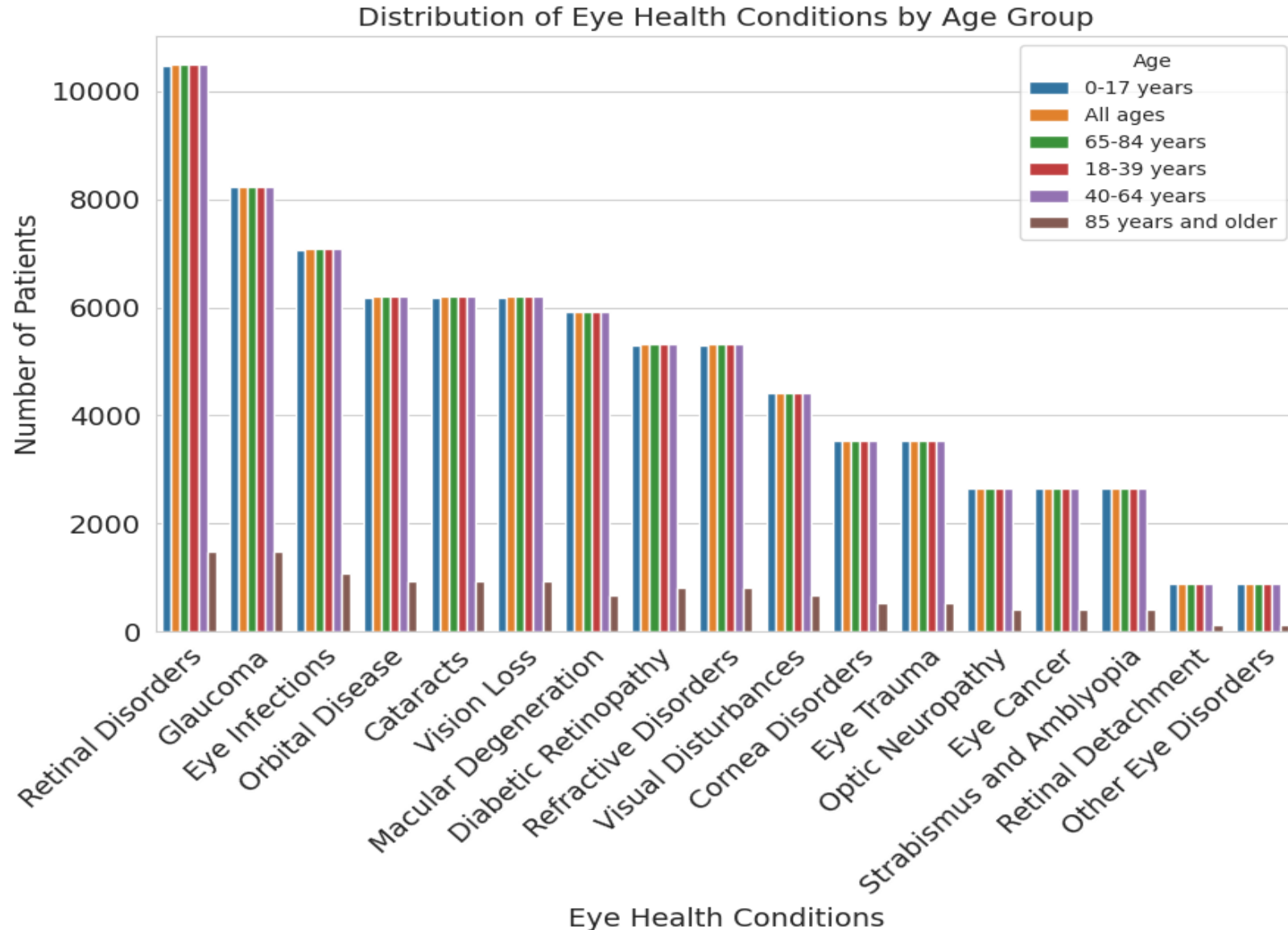
- ➢ The graph shows the relationship between different **Age Groups** and different **Eye Health** conditions.

- ➢ Except for the "85 years and older" age group, the distribution of eye disorders is **constant** across all age groups.

- ➢ Therefore, Age is not a significant factor in determining eye disorders for the selected population under review.

ANALYSIS 3 : Distribution of Eye disorders by Gender

Eye Health Conditions of Patients by Gender
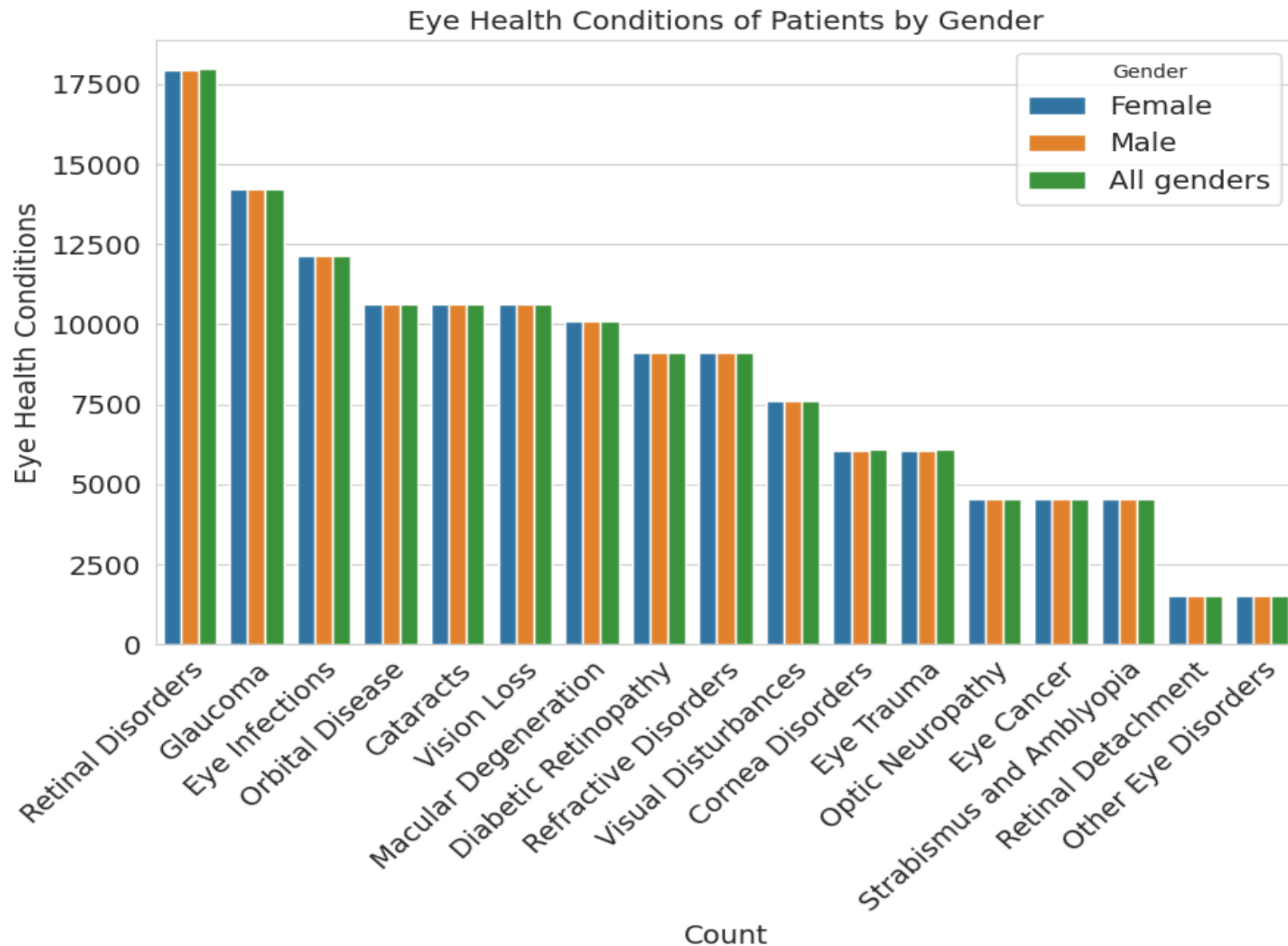
- ➤ The graph shows the relationship between different **Gender** and **Eye Health** conditions.

- ➤ The Gender distribution is the **same** across various Vision disorders, where **Retinal Disorders** contribute to the major number of people.

- ➤ Therefore, Gender is not a significant factor in determining eye disorders for the selected population under review.

- ➤ The data granularity is at the disease level rather than the individual person level, and therefore, the category "**All genders**" is present as part of the Gender column.

ANALYSIS 4 : Distribution of Risk Factors

Distribution of Risk Factors associated with the Population
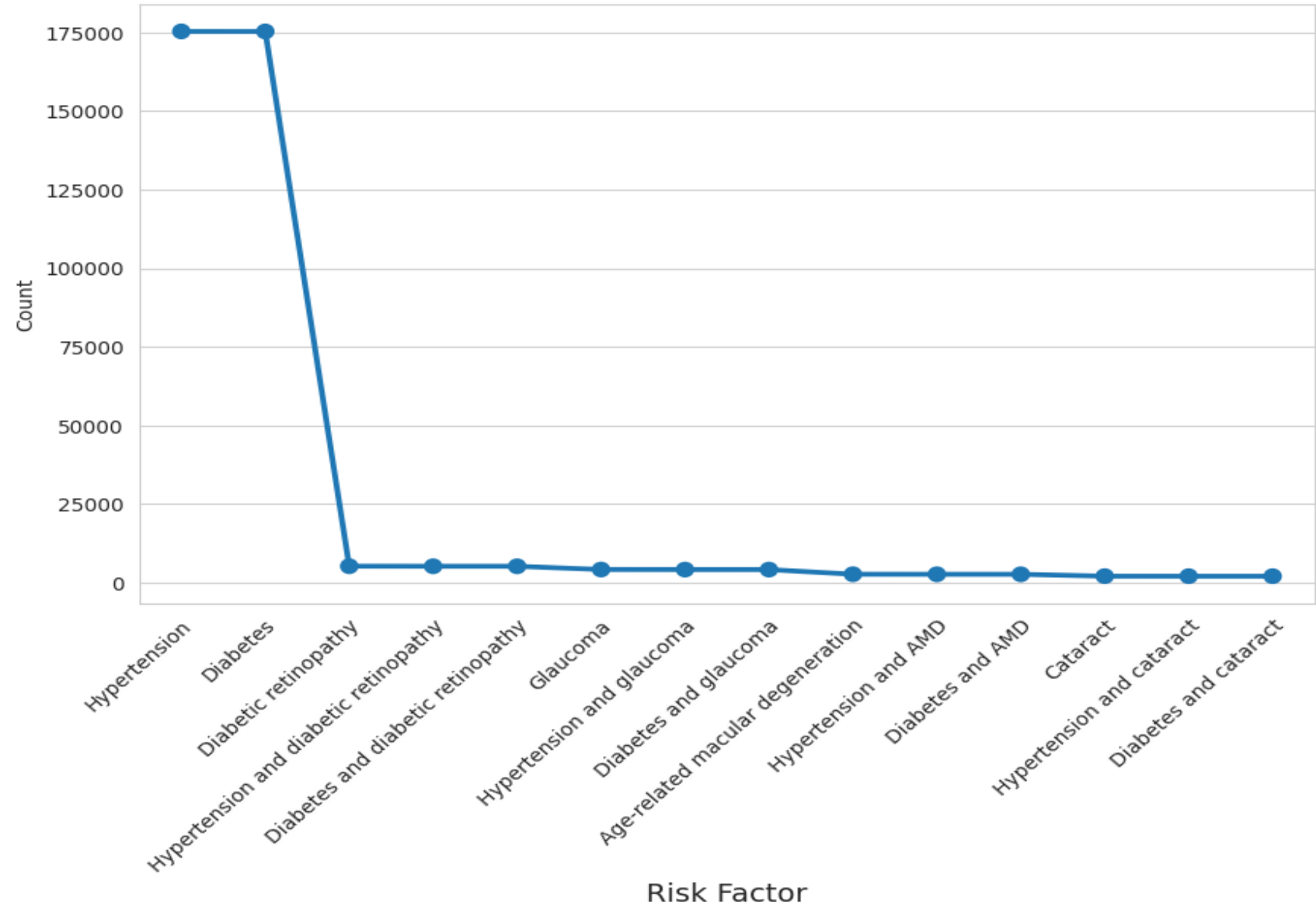
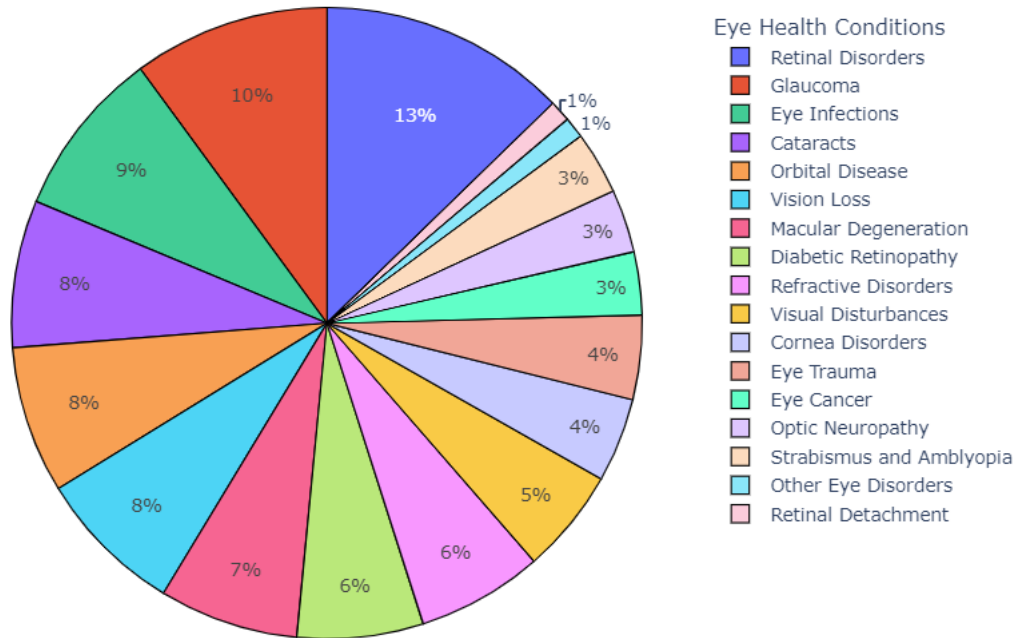The graph presents the ***Risk Factors*** associated with the population in the US.

The two major factors that stand out are ***Hypertension*** and ***Diabetes***, with values of around 17.5K each.

The other Risk Factors associated with the population are not that significant and can be ignored for analysis.
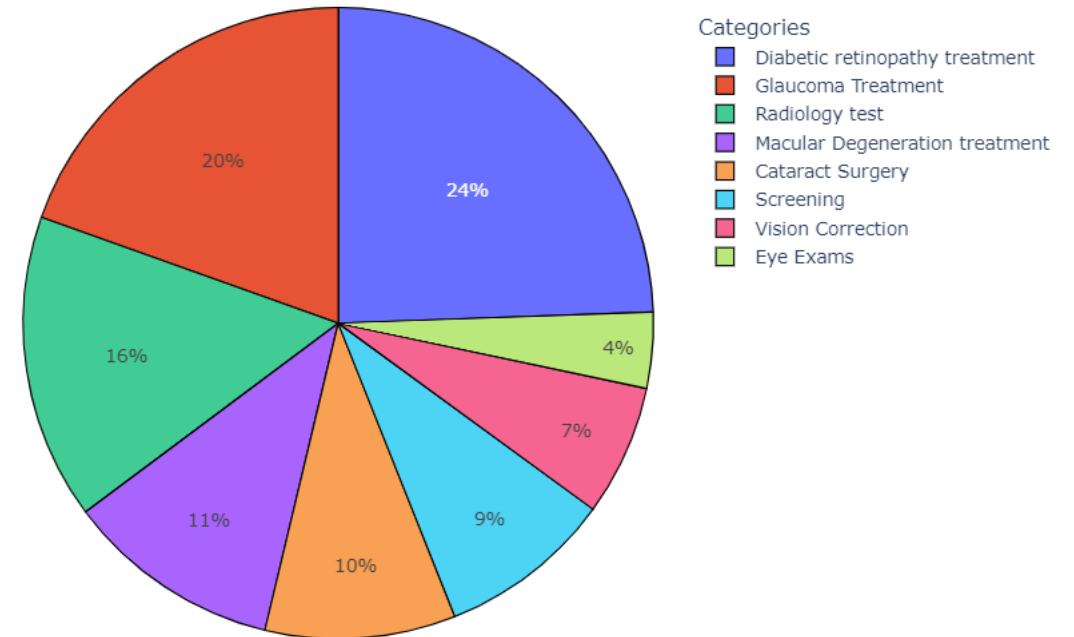
ANALYSIS 5 : Eye Disorder vs Service Utilization

Distribution of Eye Health Conditions

Eye Health Conditions
- Retinal Disorders
- Glaucoma
- Eye Infections
- Cataracts
- Orbital Disease
- Vision Loss
- Macular Degeneration
- Diabetic Retinopathy
- Refractive Disorders
- Visual Disturbances
- Cornea Disorders
- Eye Trauma
- Eye Cancer
- Optic Neuropathy
- Strabismus and Amblyopia
- Other Eye Disorders
- Retinal Detachment

Distribution of Service Utilization

Categories
- Diabetic retinopathy treatment
- Glaucoma Treatment
- Radiology test
- Macular Degeneration treatment
- Cataract Surgery
- Screening
- Vision Correction
- Eye Exams

➢ **Retinal Disorders** and **Glaucoma** together affect about _23% of the population_ suffering from vision-related issues in the US.

➢ **Diabetic Retinopathy**, **Glaucoma treatment**, and **Radiology** are the most sought-after *Medical Services* for eye health in the US.

➢ Among the Medical Services, _Diabetic Retinopathy accounted for 24%_, _Glaucoma treatment for 20%,_ and _Radiology for 16%_ of the services.

➢ The figures indicate that eye health is a significant concern in the US and that Retinal Disorders and Glaucoma are prevalent among the population.

ANALYSIS & RECOMMENDATIONS

➢ Since, **Retinal Disorders** and **Glaucoma** are prevalent among the population. Insurance companies should prioritize eye health coverage plans considering these two major issues.

➢ Among the medical services, _Diabetic Retinopathy accounted for 24%_, _Glaucoma treatment for 20%,_ and _Radiology for 16%_ of the services. Insurance companies can use this data to determine the most critical services and provide coverage accordingly to ensure that their customers receive adequate medical attention.

➢ Insurance companies should focus on _providing comprehensive coverage to all age groups and genders_ as Age and Gender do not seem to be significant factors in determining eye disorders.

➢ Based on _Risk factors_, companies may consider **higher premiums** for individuals with _Diabetes_ and _Hypertension_ due to the increased risk of eye health conditions.

➢ Thus, to be competitive in the US eye insurance market, it is imperative for insurance companies to concentrate on prevalent eye disorder categories and the most sought-after medical services.

➢ The above insights can help them design new products or enhance existing policies to drive profitability and gain an edge over their competitors.

GITHUB PROJECT REPOSITORY

https://github.com/users/rituraj-borah/projects/2/views/1

REFERENCES

Vision impairment and blindness. (2021, October 14). WHO | World Health Organization. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

Visual impairment, blindness cases in U.S. expected to double by 2050. (n.d.). National Eye Institute | National Eye Institute. https://www.nei.nih.gov/about/news-and-events/news/visual-impairment-blindness-cases-us-expected-double-2050

Open Data | Centers for Disease Control and Prevention | Chronic Disease and Health Promotion Data & Indicators. https://chronicdata.cdc.gov/Vision-Eye-Health/Commercial-Medical-Insurance-MSCANCC-Vision-and-Ey/a35h-9yn4

User guide and tutorial — Seaborn 0.12.2 documentation. https://seaborn.pydata.org/tutorial.html

Pandas isin() explained with examples. https://sparkbyexamples.com/pandas/pandas-isin-explained-with-examples/

Thank you