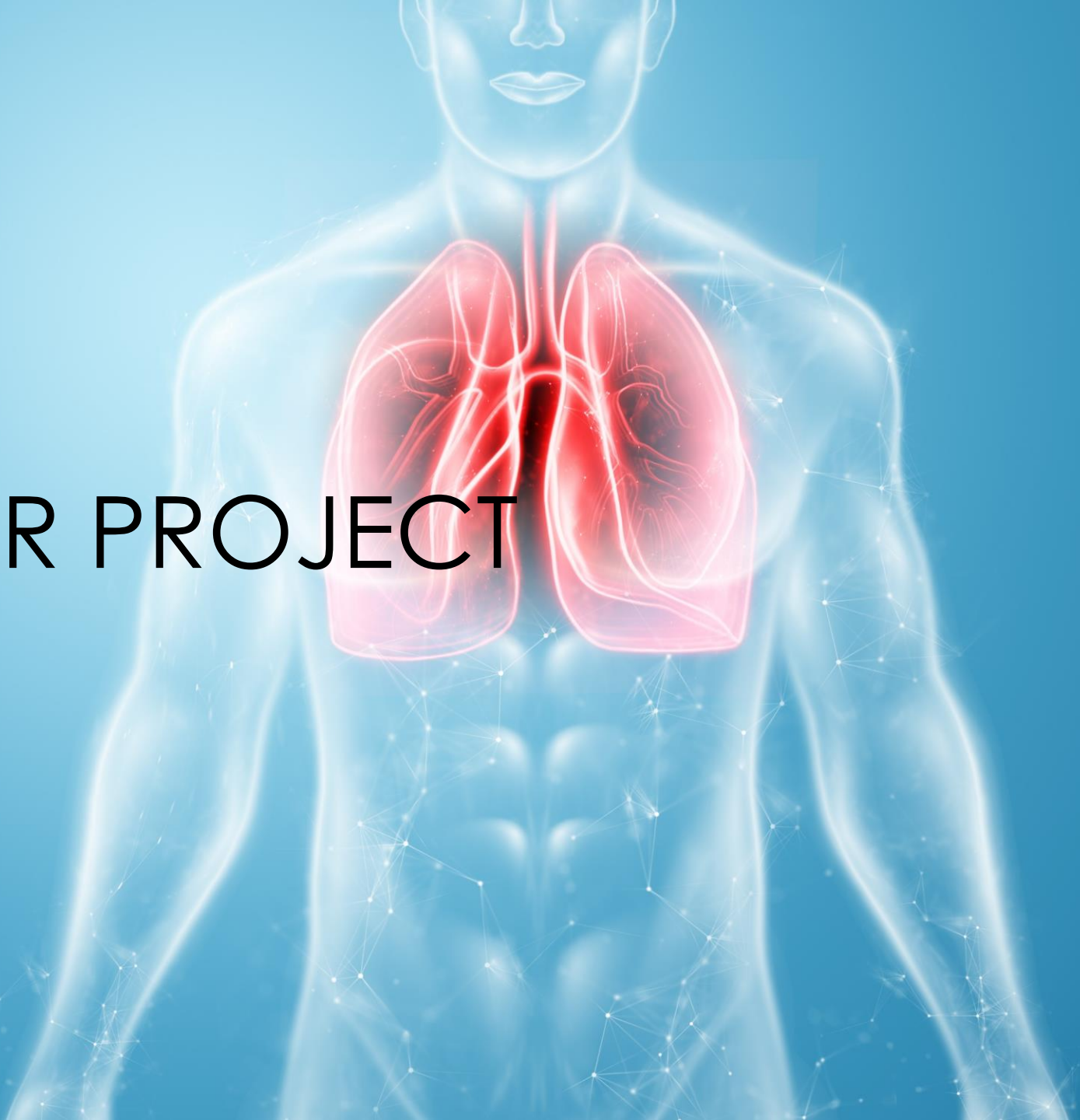


LUNG CANCER PROJECT

Using MongoDB



AIM

- ▶ The aim of this project was to use Kaggle to find a JSON dataset to upload into MongoDB and analyse.

DATASET

- ▶ As November is Lung Cancer awareness month a lung cancer database was selected for this project.
- ▶ The Dataset contains detailed patient profiles and the aim of the dataset was to allow for analysis to allow for thorough lung cancer risk assessments and prediction.
- ▶ The dataset was imported into MongoDB using Mongo Compass. The Data was then analysed in Mongo Atlas online.



Emma's Org ...



Access Manager

Billing

All Clusters

Get Help

Emma

Project 0

Data Services

Charts



Overview

DATABASE

Clusters

SERVICES

Atlas Search

Atlas Vector Search

Stream Processing

Triggers

Migration

Data Federation

SECURITY

Quickstart

Backup

Database Access

Network Access

Advanced

DATABASES: 2 COLLECTIONS: 8

VISUALIZE YOUR DATA

REFRESH

+ Create Database

Search Namespaces

Lung

Cancer.Lung.Cancer

lungcancer

sample_mflix

Lung.Cancer.Lung.Cancer

STORAGE SIZE: 208KB

LOGICAL DATA SIZE: 868.67KB

TOTAL DOCUMENTS: 3000

INDEXES TOTAL SIZE: 112KB

Find

Indexes

Schema Anti-Patterns 0

Aggregation

Search Indexes

Generate queries from natural language in Compass

INSERT DOCUMENT

Filter

Type a query: { field: 'value' }

Reset

Apply

Options

QUERY RESULTS: 1-20 OF MANY

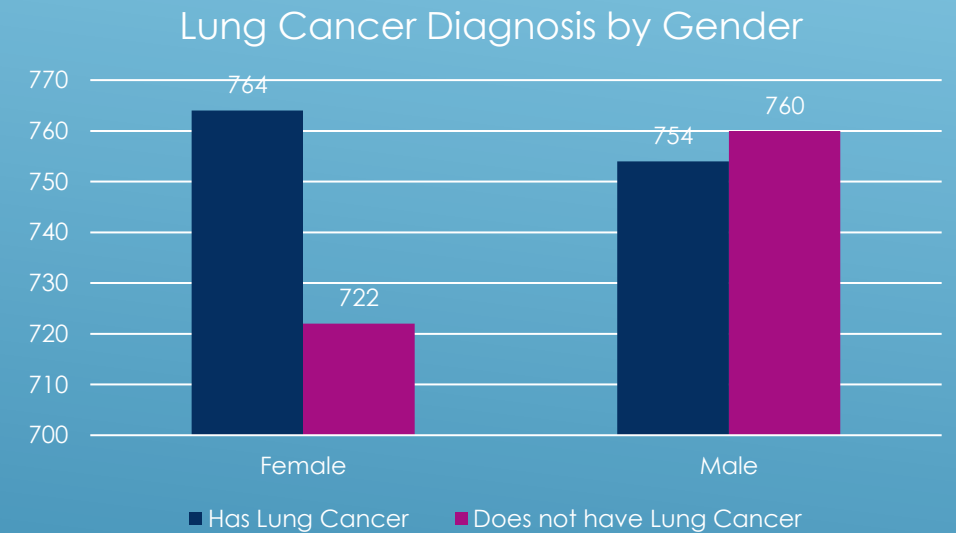
```
_id: ObjectId('67122dd69a4ee0bffc30b770')
GENDER : "M"
AGE : 65
SMOKING : 1
YELLOW_FINGERS : 1
ANXIETY : 1
PEER_PRESSURE : 2
CHRONIC_DISEASE : 2
FATIGUE : 1
ALLERGY : 2
```

SAMPLE

- ▶ The Sample Contained 1,514 Males and 1,486 Females.
- ▶ Notable categories within the data include:
 - ▶ The age at which the individual was diagnosed with lung cancer
 - ▶ The smoking status of the individual
 - ▶ Alcohol consumption of the individual
 - ▶ As well as a range of symptoms including
 - ▶ Coughing
 - ▶ Shortness of Breath
 - ▶ Chest Pain
 - ▶ Difficulty swallowing

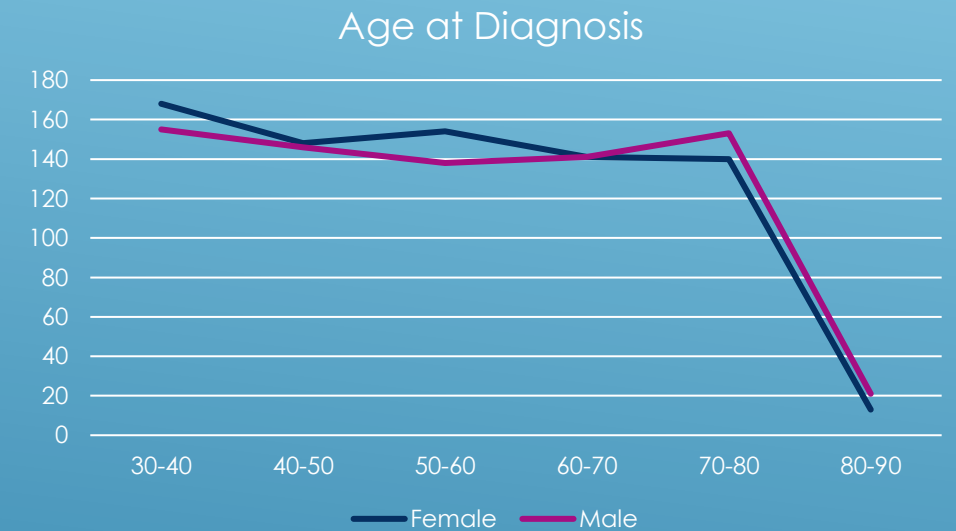
GENDER AND DIAGNOSIS

- ▶ Based on the data, slightly more women were diagnosed with lung cancer compared to men, however which is more interesting is that women were less likely to receive a negative result compared to men, although the difference is slight, with a only 5% difference.



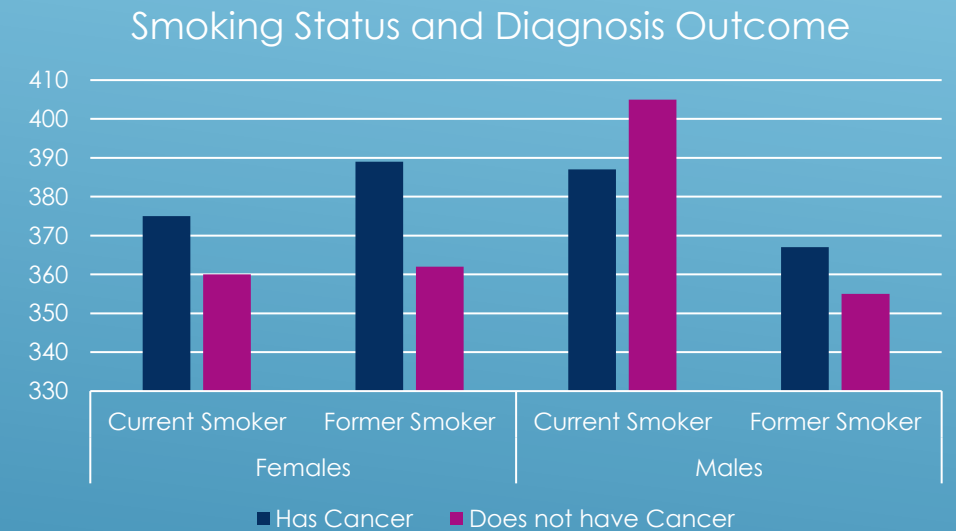
AGE AT TIME OF DIAGNOSIS

- Analysis showed that women were more likely to be diagnosed earlier than men until the age bracket of 60-70, when both groups were roughly the same. From the age of 70 onwards, more men were likely to be diagnosed.



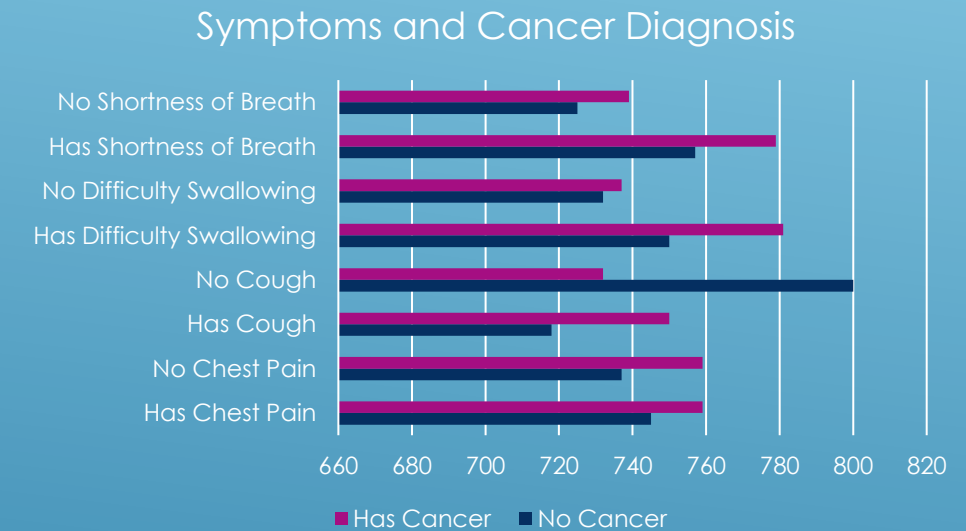
SMOKING STATUS AND DIAGNOSIS OUTCOME

- ▶ The analysis shows that within females, that whether the individual was a current or former smoker, that smoking did increase the likelihood of getting a cancer diagnosis.
- ▶ However with males, whilst former smokers confirmed to what was expected, that those who were former smokers were more likely to receive a cancer diagnosis than not, current smokers were less likely to be diagnosed with cancer, which is a surprise as this counteracts a lot of medical research on the impact of smoking and cancer.



SYMPTOMS AND DIAGNOSIS OUTCOME

- ▶ Shortness of Breath: Individuals who reported being short of breath were 2% more likely to be diagnosed with lung cancer.
- ▶ Difficulty Swallowing: Individuals who reported having difficulty swallowing were 4% more likely to be diagnosed with lung cancer.
- ▶ Cough: Individuals who reported having a cough were 4% more likely to be diagnosed with lung cancer. However 10% of people who had cancer reported having no cough at all. Suggesting that a cough is not always a strong indication of having lung cancer.
- ▶ Chest Pain: Individuals who reported having chest pain were 4% more likely to be diagnosed with lung cancer.



CONCLUSIONS

- ▶ According to this sample:
 - ▶ Females were slightly more likely to be diagnosed with lung cancer compared to men. Females were much less likely to receive a no cancer diagnosis compared to men.
 - ▶ Females were more likely to be diagnosed under the age of 60 compared to males.
 - ▶ A current or former smoking status was likely to play a factor in receiving a cancer diagnosis
 - ▶ A cough symptom is the least likely predictor of cancer diagnosis.
 - ▶ Symptoms of shortness of breath, difficulty swallowing and chest pain were more likely to produce a cancer diagnosis.

LUNG CANCER FACTS

- ▶ According to Cancer Research UK
 - ▶ Between 2017-2019 there were 49,229 new cases of lung cancer reported
 - ▶ Between 2017-2019 there were 34,771 reported death from lung cancer
 - ▶ There is a 10% chance of surviving 10 or more years with lung cancer
 - ▶ 79% of lung cancer cases are preventable.

This November, spread awareness of lung cancer and if you are worried about any symptoms you currently have, do not delay and contact your GP today. Early diagnosis is essential.