Do Clear Skies Lead to Longer Lives in Alberta?

By Jeff Foster, Carlos Arteaga, Nadeem Bishtawi, Abdul Suhaib, and Michael Moon

INTRODUCTION

Air quality is a critical determinant of public health, with long-term exposure to poor air quality linked to various adverse health outcomes. The World Health Organization (WHO) identifies particulate matter (PM), carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) as pollutants with strong empirical evidence for public health concern. Health problems can occur as a result of both short- and long-term exposure to these pollutants (World Health Organization, n.d.). Pollutants that are all part of what is considered poor air quality (*Air Quality Health Index – Calculation*, 2024b). Alberta with its diverse landscape and economic activities – including urbanization, industrial operations and natural events like wildfires` faces unique challenges in maintaining healthy air quality. Monitoring and understanding these impacts are essential for developing policies that protect public health while supporting sustainable growth.

Our project began with an exploration of Alberta's Air Quality Health Index (AHQI), aiming to understand how factors such as urbanization, industrial activity, and natural events like wildfires influences air quality. Air quality directly impacts public health, making it a critical topic for analysis. However, as we delved into the data, we encountered challenges such as inconsistencies in temporal granularity and limited data availability, especially when trying to analyze the AQHI across municipalities over an extended period. These obstacles led us to reevaluate our focus.

Through initial exploration, we found that life expectancy, another critical public health indicator, aligned with many variables we were investigating for AQHI. The shift to life expectancy as our primary focus allowed us to leverage the same datasets more effectively while addressing gaps in AQHI data. Additionally, life expectancy provided a more robust dependent variable, allowing for meaningful analysis of its relationships with demographic, environmental, and economic factors. This transition enabled us to draw connections between trends in life expectancy and variables like population growth, housing developments, and vehicle registrations, which all tie into broader societal and environmental dynamics.

In this report, we aim to explore the factors that most significantly influence life expectancy and the air quality health index in Alberta. Our analysis focuses on uncovering relationships between demographic trends, such as population growth and age distribution, and changes in life expectancy across municipalities. Additionally, we investigate how environmental factors, like wildfires, fossil fuels, and urbanization metrics may contribute to unhealthy levels of pollutants in the air that might consequently create disparities in life expectancy. By addressing these queries, we aim to provide a comprehensive understanding of the interplay between demographic, environmental, and societal factors that shape the life expectancy and air quality in Alberta.

DATASETS / METHODOLOGY

All datasets used in our investigation are collected from Alberta's open government portal (OGP). The OGP is committed to providing open access to Albertan data and publications. They actively work with ministries to retrieve accurate data. The licensing of this data is under an open government license, which can be found here. To summarize the licensing, it provides freedom to access, modify, and distribute data as long as actions are lawful and proper attributions to the source are given. Below, we have listed out datasets from the open government portal with details we have collected that we will be considering for our analysis:

DATASETS	Description	Size	Format	Variables of interest
Air Quality Index by municipality	Percentage of hrs p/year at a given air quality level.	8 cols x 2312 rows	CSV	Air quality health index, health risk, value, municipality, year.
Population by Municipality	Amount of people residing in each municipality annually	9 cols x 1,048,576 rows	CSV	Total population, municipality, year
Life Expectancy by Municipality	Average life expectancy per municipality	7 cols x 11132 rows	CSV	Life expectancy in years, municipality and year.
Housing starts by municipality	Number of housing starts as defined by the beginning of construction work on a building.	7 cols x 3491 rows	CSV	Total number of housing starts, municipality, year
Well count by municipality	Well counts including nat gas, crude oil, bitumen and others.	6 cols x 1460 rows	CSV	Well counts, municipality and year.
Natural gas production by municipality	Natural gas production by year and municipality district.	6 cols x 1647 rows	CSV	Natural gas production, municipality and year.
Greenhouse gas emissions by municipality	Annual CO2-equivalent emissions exceeding 10kt from Canadian facilities, categorized by municipality and year.	7 cols x 22356 rows	CSV	Emission type, the value of emission in kt, the municipality and year.

Historical wildfire data : 2006 to 2023	Individual wildfires in the province of Alberta including size, location, conditions, and more.	50 cols x 25,322 rows	CSV	Fire year, fire size, fire location, wind direction, wind speed.
Cattle and calves by municipality	Total number of bulls, milk & beef cows, heifers, steers, and calves on farms.	6 cols x 353 rows	CSV	Year, census subdivision, cattle and calves count.
Pigs by municipality	Number of pigs located on farms, by census year and municipal district.	6 cols x 355 rows	CSV	Year, census subdivision, pigs count.
Cropland by municipality	Total area of land dedicated to growing crops of various types, by municipal district, type of crop and census year.	7 cols x 1,707 rows	CSV	Year, census subdivision, crop type, crop acres.
Motorized Vehicle Registrations by Municipality	Number of registered vehicles in each Albertan Municipality annually by vehicle type	7 cols x 41992 rows	CSV	Year, census subdivision, Vehicle registration count
Dwelling Units by Municipality	Number of Units that can be lived as a count, ordered by municipality and year	5 cols x 7672 rows	CSV	Year, census subdivision, Dwelling unit count
Households by Structural Type by Municipality	Total number of housing units in each municipality each year, organized by structural type	7 cols x 15505 rows	CSV	Year, census subdivision, Building Type,

All of these datasets were chosen due to their potential to impact air quality. To combine these datasets, we needed to look for a commonality between all the data. Each data set selected for our investigation has common columns of municipality and year. These columns were used as keys to create our database, with municipality being our primary key and year being the secondary key.

To clean and prepare our data to be merged, we adjusted column names across all datasets to ensure the data they held would be clear to understand. We used Pandas to merge all our datasets, stacking one on top of the other, based off our keys of municipality and year. All datasets had standardized nomenclature for municipalities and years, making joining them relatively simple.

We also implemented imputations of missing municipality values for our fire data and transformed our AQHI data. We used geocoding from the OpenCage Geocoding API (OpenCage, n.d.) in Python to convert coordinates in our fire dataset to their respective municipality. Using this geocoding tool, we were able to geolocate >98% of the fires to their respective municipality.

The AQHI dataset initially contained 2,500+ rows of data, but after further exploration we realized that it had to be transformed. It originally contained data on the percentage of time at specific air quality levels on an ordinal scale of 1 to 10 (1 being the best and 10 being the worst) per year per municipality. This meant that each year per municipality contained 10 rows. The analysis required a single average value per year, and thus, data was transformed by taking the weighted average for each municipality of how long their air quality was at each respective level (1-10).

As a group, our hope was to use SQL to explore the data and attempt to leverage linear regression and machine learning techniques. We wanted to develop a predictive model based on some known independent variables to predict our dependent variable, Air Quality Index. Due to the AQHI data transformation, the size of the data set was reduced to less than 300 rows. Additionally, once the datasets were fully merged, most of the rows did not contain data across all variables, therefore, we were unable to successfully implement any predictive models (read 'Roadblock' for a more thorough explanation on this problem). Nevertheless, this AQHI average was the most meaningful insight we could obtain from the AQHI data set, and these were the values that were used in our analysis.

To mitigate these issues, we decided to shift our focus to a different dependent variable that we thought would contain more robust data. When assessing the data we already had collected, we agreed as a team that life expectancy would be the most logical next dependent variable to analyze. While our life expectancy data was more robust and had some relations with other data already collected in our database, in the end, we ran into the same problem as before, where data was not sufficient and/or meaningful across all variables. Finally, due to our late transition in topics and time constraints of the project and since AQHI and life expectancy are both health-related issues, we decided to explore both topics against the data already collected.

Our research questions all revolve around the impact of the different demographic and environmental factors on life expectancy and/or air quality, and the interaction of both. Each question can be found as a sub header in the analysis section.

ROADBLOCK: STATISTICAL MODELLING

As discussed, from the inception of our project, our intention was to create a multivariate regression analysis using AQHI as our dependent variable on a variety of independent variables gathered in our datasets. This model would allow us to understand what were the key variables of prediction and allow us to predict the resulting Air Quality based on the known independent variable. With this model created, our continued plan was to create a machine learning model that would allow us to predict values moving into future years. Unfortunately, due to significant data challenges, we were able to move past the statistical model testing.

Our first problem was that the initial AQHI data set contained over 3,000 records, which, once transformed to represent the average AQHI per municipality per year, went down to less than 300 records. An initial temp to fit a model with all potential predictors is displayed below:

```
Residuals:
ALL 9 residuals are 0: no residual degrees of freedom!

Coefficients: (16 not defined because of singularities)
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                      -1.036e+03        NaN     NaN      NaN
CSDCalgary                        3.516e+01        NaN     NaN      NaN
CSDEdmonton                       2.696e+01        NaN     NaN      NaN
CSDFoothills County               1.120e+01        NaN     NaN      NaN
Period                            5.832e-01        NaN     NaN      NaN
TOTAL.POPULATION                 -2.174e-05        NaN     NaN      NaN
Life.Expectancy..Years.          -1.815e+00        NaN     NaN      NaN
Oil.Wells..Count.                -1.333e-04        NaN     NaN      NaN
Natural.Gas.Production..m.3.      5.754e-06        NaN     NaN      NaN
CH4.Equivalent..t.                       NA         NA      NA       NA
CO2.Equivalent..t.                       NA         NA      NA       NA
N2O.Equivalent..t.                       NA         NA      NA       NA
Oats..Count.                             NA         NA      NA       NA
Canola..Count.                           NA         NA      NA       NA
Barley..Count.                           NA         NA      NA       NA
Durum.Wheat..Count.                      NA         NA      NA       NA
Non.Durum.Wheat..Count.                  NA         NA      NA       NA
Other.Crops..Count.                      NA         NA      NA       NA
Cattle.and.Calves..Count.                NA         NA      NA       NA
Pigs..Count.                             NA         NA      NA       NA
Dwelling.Units..Count.                   NA         NA      NA       NA
Residence.Housing.Starts..Count.         NA         NA      NA       NA
Vehicle.Registration..Count.             NA         NA      NA       NA
Fire.Size..Hectares.                     NA         NA      NA       NA
Fires..Count.                            NA         NA      NA       NA

Residual standard error: NaN on 0 degrees of freedom
  (275 observations deleted due to missingness)
Multiple R-squared:      1,	Adjusted R-squared:    NaN 
F-statistic:   NaN on 8 and 0 DF,  p-value: NA
```

A key insight from this model are the 275 observations which were deleted due to missing data across the variables, thus resulting in a Null model. Subsequently, we started removing independent terms to try to find a model that at least provided coefficient values with a threshold of an alpha value of 0.05. An attempt using stepwise regression methodologies was performed, and the output was a statistically insignificant simple linear regression model with the dependent variable AQHI and the sole predictor variable, ‘other crops’. After lengthy sequential elimination regression and selecting the most appropriate independent variables that might be closely related to AQHI, we were able to come up with the model demonstrated below:

```
Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -1.969e+01  2.598e+01  -0.758    0.483
CSDCalgary                      1.011e+02  1.118e+02   0.905    0.407
CSDEdmonton                     7.693e+01  9.139e+01   0.842    0.438
CSDFoothills County             4.545e-01  1.108e+00   0.410    0.699
CSDYellowhead County            1.677e+00  1.415e+00   1.185    0.289
TOTAL.POPULATION                1.055e-05  2.086e-04   0.051    0.962
Life.Expectancy..Years.         2.771e-01  3.253e-01   0.852    0.433
Oil.Wells..Count.               3.672e-04  2.862e-04   1.283    0.256
Natural.Gas.Production..m.3.   -4.248e-08  1.196e-07  -0.355    0.737
Cattle.and.Calves..Count.      -2.976e-06  9.499e-06  -0.313    0.767
Residence.Housing.Starts..Count. -2.165e-03  2.657e-03  -0.815    0.452
Vehicle.Registration..Count.   -8.953e-05  1.893e-04  -0.473    0.656
Fire.Size..Hectares.            3.645e-04  2.100e-04   1.736    0.143
Fires..Count.                  -7.776e-02  7.877e-02  -0.987    0.369

Residual standard error: 0.1281 on 5 degrees of freedom
  (265 observations deleted due to missingness)
Multiple R-squared:  0.9508,    Adjusted R-squared:  0.8229
F-statistic: 7.433 on 13 and 5 DF,  p-value: 0.0186
```

While this method produced a model with coefficients, 265 records were removed due to missing data, leaving us with only 5 degrees of freedom (DF). The DF are the number of observations that remain after accounting for the parameters that we are estimating in the model, and it is computed by taking the number of observations minus the number of parameters. With only 13 parameters + the intercept and after removing the 265 observations, that means we only have 5 data points remaining to perform the regression on, and this is indicative of extreme overfitting.

Our next step was to reduce the model even more to improve our DF. After manually removing the insignificant terms, a model was created with only two terms and 122 DF. While both terms are statistically significant, the adjusted R squared is only 0.23, meaning that our model explains ~23% of the variance in the AQHI variable. With a low adjusted R squared, few significant variables, and significant overfitting concerns, we concluded that the AQHI variable is not data to fit a regression model.

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.68226 -0.09241  0.01366  0.11845  0.46577 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         2.114e+00  2.112e-02 100.073  < 2e-16 ***
TOTAL.POPULATION    8.448e-07  1.486e-07   5.683 9.18e-08 ***
Oil.Wells..Count. -1.228e-04  4.527e-05  -2.712  0.00766 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2023 on 122 degrees of freedom
  (159 observations deleted due to missingness)
Multiple R-squared:  0.2451,	Adjusted R-squared:  0.2328 
F-statistic: 19.81 on 2 and 122 DF,  p-value: 3.55e-08
```

In an attempt to salvage our regression and machine learning plans, we shifted the focus of our project to life expectancy, with our goal being to create an improved model using life expectancy as the dependent variable. Unfortunately, we came across similar challenges.

```
Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    1.537e+02  3.672e+02   0.419    0.748
CSDCalgary                    -7.769e+01  2.693e+01  -2.885    0.212
CSDEdmonton                   -3.728e+01  1.494e+01  -2.496    0.243
CSDFoothills County            5.881e+00  1.858e+00   3.165    0.195
CSDWood Buffalo               -1.583e+00  2.014e+00  -0.786    0.576
Period                        -3.523e-02  1.819e-01  -0.194    0.878
Total.Population              -9.636e-04  3.100e-04  -3.108    0.198
Oil.Wells..Count.              3.869e-04  3.792e-04   1.020    0.494
Natural.Gas.Production..m.3.  -1.079e-06  2.121e-06  -0.509    0.701
CH4.Equivalent..t.            -1.032e-06  3.745e-07  -2.755    0.222
CO2.Equivalent..t.            -3.256e-08  1.015e-07  -0.321    0.802
N2O.Equivalent..t.            -1.480e-06  1.748e-06  -0.847    0.553
Dwelling.Units..Count.         8.132e-04  2.952e-04   2.755    0.222
Residence.Housing.Starts..Count. -9.448e-05  2.661e-04  -0.355    0.783
Vehicle.Registration..Count.   6.506e-04  2.611e-04   2.492    0.243
Fire.Size..Hectares.          -2.276e-06  1.491e-06  -1.526    0.369
Fires..Count.                  6.755e-02  2.741e-02   2.464    0.245
Adjusted_AQHI                  1.698e-01  5.467e-01   0.311    0.808

Residual standard error: 0.1392 on 1 degrees of freedom
  (3697 observations deleted due to missingness)
Multiple R-squared:  0.9995,	Adjusted R-squared:  0.9911 
F-statistic:   119 on 17 and 1 DF,  p-value: 0.07198
```

Similar to our initial model using the AQHI variable as our dependent variable, our initial model using average life expectancy as our dependent variable removed 3,697 data points due to missingness, leaving us with 1 degree of freedom. Furthermore, all of our independent variables were statistically insignificant (less than 0.05). As a last-ditch effort to create a reasonable model, we attempted to follow the same sequential reduction process as previously mentioned. After sequentially removing the independent variables with the highest p values, the model displayed below was created:

```
Call:
lm(formula = Life.Expectancy..Years. ~ Total.Population + CO2.Equivalent..t. +
    Vehicle.Registration..Count., data = life_exp)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4873 -1.0223 -0.0612  1.0802  7.3377

Coefficients:
                               Estimate Std. Error  t value Pr(>|t|)
(Intercept)                   7.980e+01  3.665e-02 2177.166  < 2e-16 ***
Total.Population             -2.090e-05  7.520e-06   -2.779  0.00548 **
CO2.Equivalent..t.            6.741e-08  1.180e-08    5.711 1.23e-08 ***
Vehicle.Registration..Count.  2.473e-05  7.689e-06    3.217  0.00131 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.892 on 3036 degrees of freedom
  (676 observations deleted due to missingness)
Multiple R-squared:  0.03187,   Adjusted R-squared:  0.03091
F-statistic: 33.31 on 3 and 3036 DF,  p-value: < 2.2e-16
```

While the model is valid with sufficient sample size (3036 DF) with statistically significant independent variables, an adjusted R squared value of 0.03 informs us that our model does not accurately predict the average life expectancy. Therefore, we concluded that our data does not contain sufficient quantities of quality information for us to be able to perform multivariate regression and machine learning models. The analysis of our datasets is limited to data exploration queries and visuals.

DATA EXPLORATION

WILDFIRES

During the exploration of the wildfires data set in relation to air quality and life expectancy, we identified two columns of interest that would directly relate to our guiding questions. We focused on the fire count and the fire size (or total area burnt) and decided to start looking into municipal data, and later on, aggregating the results to hone in on province-wide results

Q. What were the worst years for Albertan municipalities when it comes to wildfires?

To tackle this question, we started with a query for the worst year in fire per municipality from 2012 to 2023:

```
fcount = pd.read_sql_query('''
SELECT firec.muni, firec.year, firec.fire_count
FROM master firec
INNER JOIN
    (SELECT muni, MAX(fire_count) AS max_fire_count
    FROM master WHERE year > 2011
    GROUP BY muni) max_fires
ON firec.muni = max_fires.muni AND firec.fire_count = max_fires.max_fire_count
WHERE firec.fire_count > 0
ORDER BY firec.fire_count DESC;''' , engine)
print(fcount)
```

And the same for the worst year in regard to fire size:

```
fsize = pd.read_sql_query('''
SELECT firea.muni, firea.year, firea.fire_size
FROM master firea
INNER JOIN
    (SELECT muni, MAX(fire_size) AS max_fire_size
    FROM master WHERE year > 2011
    GROUP BY muni) max_size
ON firea.muni = max_size.muni AND firea.fire_size = max_size.max_fire_size
WHERE firea.fire_size > 0
ORDER BY firea.fire_size DESC;''' , engine)
print(fsize)
```

These two queries revealed alarming information: the municipality of Mackenzie County appears at the top of the list for both the highest fire count at 55 in 2015 and the largest area burnt in 2023 at 577,721.79 hectares. Wood Buffalo was second in both lists with 44 fires in 2022 and 497,393.98 hectares burnt in 2016. As expected, most of the same municipalities appear on both lists, but most worrying is the fact that the highest fire counts and fire sizes occurred in the last 8 years, and especially 2023, the last year that we have data for.

Taking a step back and looking at the whole picture, the following query shows that 2023 has been the worst year by far in the last 11 years with a total area of 1,661,252.01 hectares burnt, distantly followed by 2016 with 505,256.80, which only represents 30% of the total area burnt in 2023.

```python
ftotal = pd.read_sql_query('''
SELECT year, SUM(fire_count) AS total_fire_count, SUM(fire_size) AS total_area_burnt
FROM master
WHERE year > 2011
GROUP BY year
ORDER BY year;''', engine)
print(ftotal)
```

Looking at the wildfire trend plot below, we see that fire size and fire count aren't necessarily correlated. While there are discrepancies around the major cause for fires, whether it's lightning or human carelessness, most researchers and scientists agree that climate change is driving this problem to higher levels breaking records in Canada in 2023 (Drew, 2023).

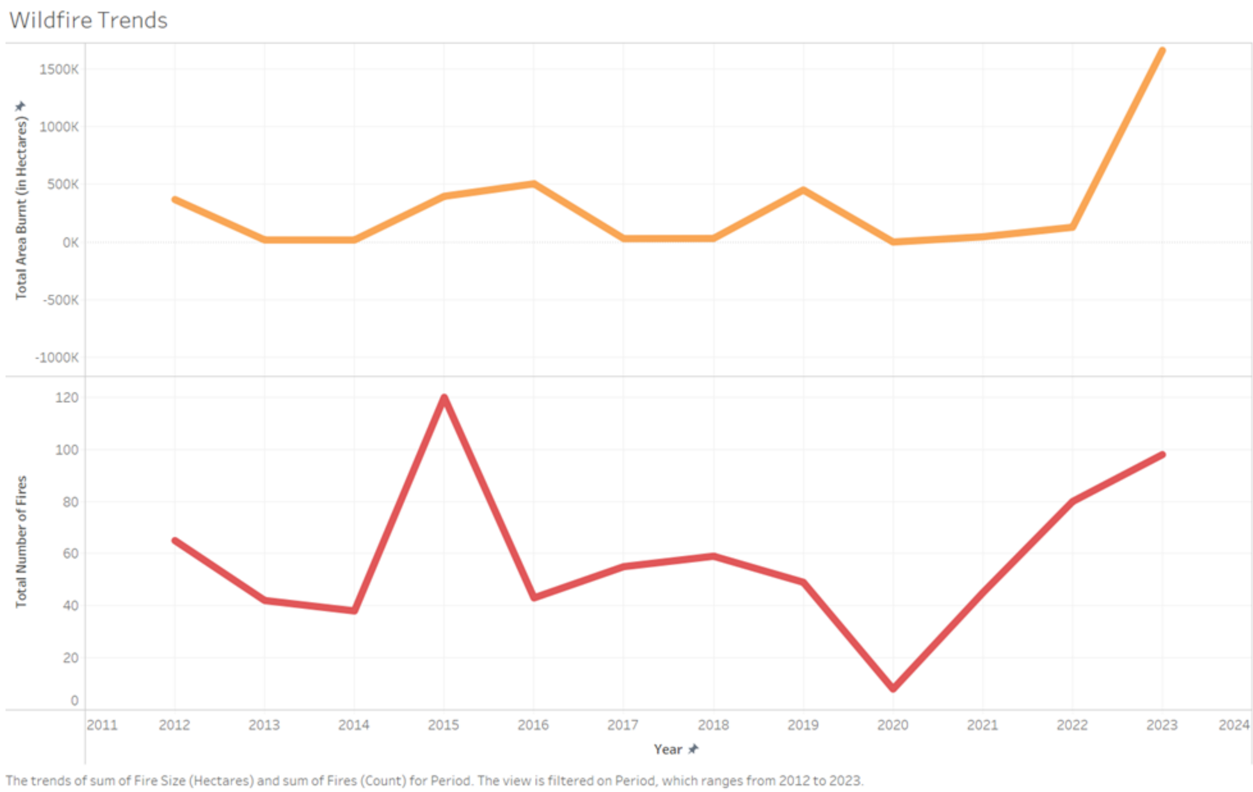

Figure 1: Comparing trends between wildfire count and fire size in hectares.

Q. What is the impact of wildfires on air quality and life expectancy?

Besides the obvious environmental impact of wildfires, in the most recent years we have witnessed a significant increase in bad air quality during fire season. Adding to this, the rising global temperatures have contributed to the "...phenomenon of winds pushing smoke down..." to the southern parts of Canada and northern U.S. (Kopeliovich, 2023), carrying the smoke from wildfires that are kilometers away to densely populated areas, threatening people's respiratory and hematologic health (*Wildfire Smoke and Your Health*, n.d.).

Calgary is the most densely populated municipality in Alberta, and it turned out to have the worse average AQHI in the province. With the purpose to discover what period was the worst, the following query revealed that the top 10 worst years for air quality for most municipalities was 2018 even though there were no wildfires in any of them during the same year. This might be attributed to wind speed and direction, since the wind can be blowing smoke from other provinces and from fires far away, for example in BC, which had its second worst year for wildfires in 2018 (BC Wildfire Service, 2024).

```
fires_vs_aqhi = pd.read_sql_query('''
SELECT ot.muni, ROUND(ot.aqhi, 3) AS highest_aqhi, ot.year, ot.fire_count
FROM master ot
JOIN
    (SELECT muni, MAX(aqhi) AS max_aqhi
    FROM master
    WHERE aqhi != ''
    GROUP BY muni
    HAVING COUNT(year) > 4) inn
ON ot.muni = inn.muni AND ot.aqhi = inn.max_aqhi
WHERE ot.aqhi != ''
ORDER BY ROUND(ot.aqhi, 3) DESC
LIMIT 10;
''', engine)

print(fires_vs_aqhi)
```

During fire season we mostly hear about citizen displacement, and fortunately not a lot about deaths. Regardless, we were interested to see if we could find any trends between life expectancy in the province and wildfires. As expected, we didn't notice any trends. On the one hand, a huge subset of the population would have to die for this to be reflected in life expectancy, and on the other, the effects of smoke on health, are not apparent immediately. Moreover, this query provided some unexpected results. We discovered that life expectancy peaked in 2019 at 80.17 years-old, and it has been going down since then. It is hard to speculate about why this is, since there are countless factors that have an impact on the life-span of Albertans.

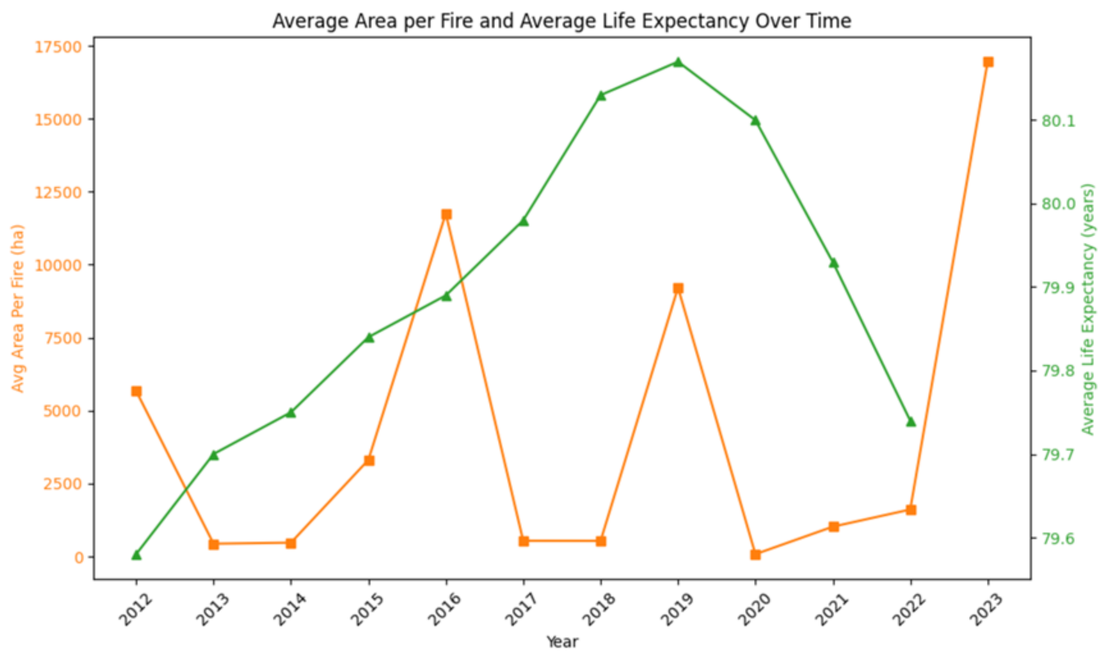

Figure 2: Life expectancy VS average fire size (in hectares)

ENERGY

The energy sector constitutes a pillar of every society and as much as it drives innovation and progress, fossil fuels come at a cost to the environment and consequently health. For our comparative analysis, we decided to hone in on greenhouse gas emissions compared to natural gas production and oil wells around the province.

Q. How does Alberta's energy sector impact the air quality health index?

Alberta's oil sands are mostly located in the northeast of the province, and expectedly, municipalities located in that region tend to have higher levels of energy production.

```
df_query_3 = pd.read_sql_query('''
    SELECT
        CSD,
        (SUM(`CO2 Equivalent (t)`) + SUM(`N2O Equivalent (t)`) + SUM(`CH4 Equivalent (t)`)) * 100.0 /
        (SELECT SUM(`CO2 Equivalent (t)`) + SUM(`N2O Equivalent (t)`) + SUM(`CH4 Equivalent (t)`) FROM masterdata)
        AS contribution_percentage,
        AVG(`Adjusted_AQHI`) AS average_aqhi
    FROM
        masterdata
    GROUP BY
        CSD
    ORDER BY
        contribution_percentage DESC
    LIMIT 5;
''', engine)
```

This first query demonstrates the contribution of municipalities of total emission in Alberta, compares to their AQHI. The municipality of Wood Buffalo, located in the heart of the oil sands, has the highest emissions contribution at 29.5% but a great AQHI yearly average of 1.98. The rest of the municipalities show similar trends with no direct correlation between average AQHI and emissions. Since 1987 there has been an emphasis of maintaining good air quality in Wood Buffalo where the emissions per gallon of oil have been reduced since 1990 ("Alberta's Oil Sands Resourceful. Responsible.," 9 C.E.). Nevertheless, this municipality is still a major source of greenhouse gas emissions, and a more thorough investigation would be insightful to determine what exactly is being done to reduce the impact on AQHI.

Much has been discussed about phasing out fossil fuels in Alberta and analyzing oil well trends paint a less bleak picture.

```
totaloil2= pd.read_sql_query('''
SELECT year, SUM(well_count) AS total_oil_wells, ROUND(AVG(aqhi), 3) AS avg_aqhi
FROM master
WHERE well_count != ''
GROUP by year
ORDER BY year;''', engine)
print(totaloil2)
```

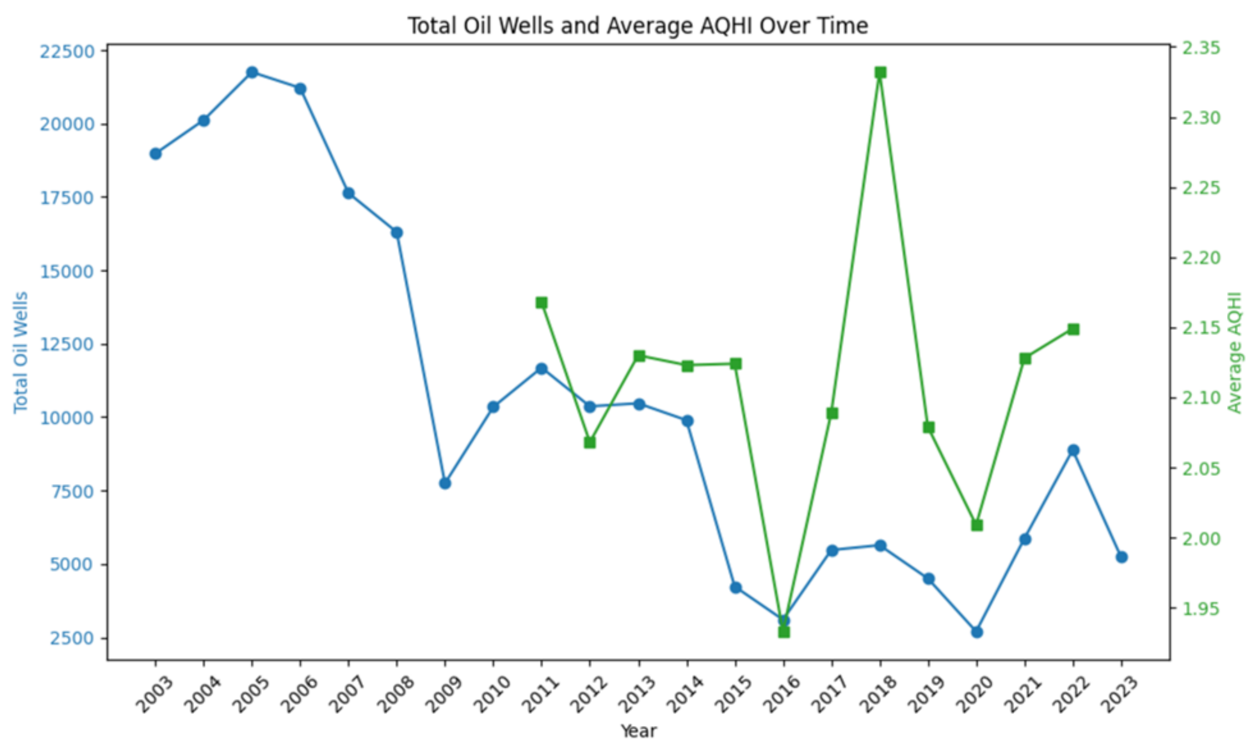

FIGURE 3: Trends in number of oil wells across Alberta compared to average AQHI

As we can see, oil wells peaked in 2005 with a total of 21,75 oil wells in the province compared to the lowest in 2020 with only 2,699. Like greenhouse gas emissions, we couldn't see any correlation between oil well decline and an improvement in air quality. Nevertheless, we wanted to dive a bit deeper into the municipalities that have a higher-than-average oil well count.

```
oil_avg_aqhi = pd.read_sql_query('''
SELECT oilw.muni, oilw.year, oilw.well_count, inner_table.avg_count, ROUND(oilw.aqhi,2) AS aqhi, inner_table2.avg_aqhi
FROM master AS oilw
JOIN
    (SELECT year, ROUND(AVG(well_count),2) AS avg_count
     FROM master
     WHERE well_count != ''
     GROUP BY year) inner_table
ON oilw.year = inner_table.year
JOIN
     (SELECT year, ROUND(AVG(aqhi),2) AS avg_aqhi
     FROM master
     WHERE aqhi != ''
     GROUP BY year) inner_table2
ON oilw.year = inner_table2.year
WHERE oilw.well_count > inner_table.avg_count
AND oilw.aqhi != ''
ORDER BY oilw.year;
''', engine)
print(oil_avg_aqhi)
```

This last query was the definitive indicator that there is indeed no apparent correlation between the two. While the same municipalities kept showing up year after year for a higher-than-average well count (Wood Buffalo, Yellowhead County, Grande Prairie, etc.), the average AQHI per year for almost all of these municipalities was very close to the average AQHI of municipalities with little to no oil wells.

In accordance with our data, even though Alberta is the largest greenhouse gas emitter in the country (Government of Canada, Canada Energy Regulator, 2023), the impacts on AQHI don't seem to be significant, and consequently, we don't see how this would have an effect on life expectancy.

INANIMATE DEMOGRAPHICS

To relate air quality and life expectancy to inanimate demographics in Alberta, we investigated key contributors of air pollution including vehicle registration, dwelling counts, and housing starts. Our guiding questions on this topic were all led by how these factors may affect air quality and/or life expectancy. We first explored how inanimate factors relate to air pollution as shown by our data. We then ran into queries using air quality index as our response variable, and will show how we shifted to exploring life expectancy against these inanimate factors.

Q. What is the relationship between inanimate demographics and air pollution?

```
df_query_5 = pd.read_sql_query('''
    SELECT
        CSD,
        SUM(`Dwelling Units (Count)`) AS total_dwelling_units,
        SUM(`Vehicle Registration (Count)`) AS total_vehicle_registrations,
        (SUM(`CO2 Equivalent (t)`) + SUM(`N2O Equivalent (t)`) + SUM(`CH4 Equivalent (t)`)) * 100.0 /
        (SELECT SUM(`CO2 Equivalent (t)`) + SUM(`N2O Equivalent (t)`) + SUM(`CH4 Equivalent (t)`) FROM masterdata)
        AS total_emissions_percentage,
        SUM(`Dwelling Units (Count)`) - SUM(`Vehicle Registration (Count)`) AS dwelling_vs_vehicle_diff
    FROM
        masterdata
    GROUP BY
        CSD
    ORDER BY
        total_vehicle_registrations DESC
    LIMIT 5;
''', engine)
```

For all the municipalities in our datasets, the number of vehicle registrations significantly exceeds the number of dwelling units. Strathcona County shows a strong relationship between higher vehicle ownership and emissions contribution. Despite fewer total vehicles (1.37M) compared to urban centers, its emissions percentage (6.51%) is disproportionately high. This could very well be due to other emission contributing factors possibly linked to oil and gas production. Urban areas like Calgary and Edmonton have the largest differences between dwelling units and vehicle registration, with 4.17M and 2.02M more vehicles than dwellings, but while Calgary has a 3.6% contribution to emissions, Edmonton only contributes with 1.5%. These insights indicate that there is a higher per-household vehicle ownership, or a great number of vehicles registered in these cities while not necessarily living in them. Lastly, the lack of strong correlation between high vehicle registration and dwelling units with larger percentage emissions, would point back to oil and gas as the major source of greenhouse gas emissions.

Q. Can we identify any patterns in the data that give us insight on how inanimate factors affect AQI / Life Expectancy?

```
query5 = pd.read_sql_query("""
    SELECT CSD,
        AVG(Total_Population)
            AS mean_total_population,
        AVG(`Dwelling Units (Count)`) / AVG(Total_Population)
            AS dwelling_to_population_ratio,
        AVG(`Vehicle Registration (Count)`) / AVG(Total_Population)
            AS vehicle_registration_to_population_ratio,
        AVG(`Residence/Housing Starts (Count)`) / AVG(Total_Population)
            AS housing_starts_to_population_ratio,
        AVG(`Life Expectancy (Years)`)
            AS mean_life_expectancy,
        AVG(Adjusted_AQHI)
            AS mean_aqi
    FROM masterdata
    GROUP BY CSD
    ORDER BY CSD
    """, engine)

print(query5)
```

To answer this question, we developed a query (as shown above) to investigate our demographic factors collected relative to population size. Population is a key metric when considering these variables, because both vehicle registration and dwelling count are relative to the population size of a municipality. Because of this, we took the ratio for vehicle registration and dwelling count against population size. Creating visualizations from this query was able to give us more perspective on the results:

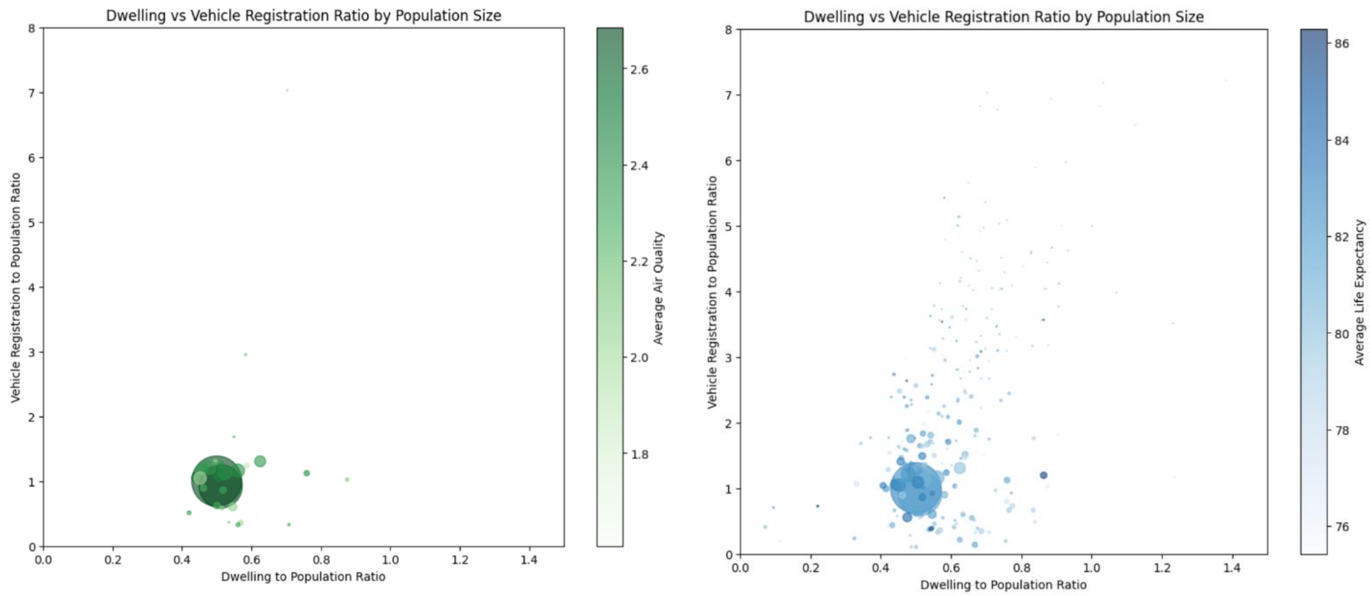

Figure 4: Bubble plots representing how dwelling count and vehicle registration rates per person compare. The size of each bubble is representative of population size, and the color schematic is representative of AQI (left) and life expectancy (right).

As we can see from the above visualizations, our initial attempt to use this data against the air quality index did not yield robust results. Since our transformation drastically reduced the amount of data we were working with, and our air quality index data is for limited municipalities, there was not enough data points to draw many conclusions on how dwelling count and vehicle registration affect air quality. Nevertheless, we can see that there are no obvious trends in how life expectancy is being affected by either vehicle registration or dwelling count, which is to be expected. The insights we can take from the graphic relate more to how the demographic variables relate to one another and to population size. We can see that for most municipalities in Alberta, there is a ratio of ~2:1 for population: dwelling count and there is a ratio of ~1:1 for population: vehicle registration.

One interesting fluctuation from normal is the many datapoints where vehicle registration is higher than two come from small municipalities. It makes sense that people own more vehicles in smaller communities, as there is less walkability and access to public transportation. Vehicles like tractors used by farmers are not included in this data, as they are not registered under motor vehicles (*Alberta Regulatory Info*, n.d.). This means there are likely even more vehicles in some smaller communities per dwelling unit that are not represented here.

Q. How are housing starts and population growth rates correlated?

If we are interested in how population growth and the number of houses/residences being built has grown in Alberta, we mostly have to zoom in on Calgary and Edmonton, sine these two cities are significantly raising the average compared to the rest of the province that has significantly smaller populations (except for Lethbridge and Red Deer).

```
yyc_yeg_pop = pd.read_sql_query('''
SELECT muni, year, total_population AS tot_pop,
       ROUND(
           CASE
               WHEN muni = 'Calgary' THEN (total_population - 741043) / 741043 * 100
               WHEN muni = 'Edmonton' THEN (total_population - 542126) / 542126 * 100
               ELSE 0
           END, 2) AS pop_g_rate,
           housing_starts,
           ROUND
           (
           CASE
               WHEN muni = 'Calgary' THEN (housing_starts - 11349) / 11349 * 100
               WHEN muni = 'Edmonton' THEN (housing_starts - 7855) / 7855 * 100
               ELSE 0
           END, 2) AS housing_g_rate
FROM master
WHERE muni IN ('Calgary', 'Edmonton')
AND total_population != '' AND housing_starts !=''
ORDER BY year;
''', engine)

print(yyc_yeg_pop)
```

The results of the query shown above, demonstrate that Calgary's population has expanded by 49.4% compared to Edmonton's 56% increase from 2001 to 2023. The housing industry shows a different trend where Calgary's housing starts have grown by 72.5% compared to Edmonton's 67.8% in the same period. Despite housing seemingly growing faster than the population, there are several factors that would tell a different story, such as housing affordability, housing maintenance and the shift in living habits in the last 20 – 30 years (younger people becoming independent and moving out of their parent's house at an earlier age). Additionally, housing starts have been very inconsistent with periods of growth and decline over the years while population has always maintained a steady growth. In the end, the best indicator of the relation between population and housing availability would be the current housing market that is constantly on the news telling us that the number of residences being built are not enough (or affordable) compared to the population growth.

ANIMATE DEMOGRAPHICS

To explore the relationship between air quality, life expectancy and animate demographics in Alberta, we focused on variables such as population growth and age distribution. Our guiding questions aimed to understand how these demographic factors may influence air quality or life expectancy trends across municipalities.

Q. How does the population growth rate impact life expectancy trends across Alberta's municipalities?

While population size offers insights into life expectancy disparities, analyzing population growth rates provides a dynamic perspective on how demographic changes impact health outcomes. Using our merged dataset, we calculated year-over-year growth rates and paired them with changes in life expectancy across Alberta's municipalities.

```
query_top_results = """
SELECT
    CSD,
    Period,
    ROUND(((Total_Population - LAG(Total_Population) OVER (PARTITION BY CSD ORDER BY Period)) * 1.0
    / LAG(Total_Population) OVER (PARTITION BY CSD ORDER BY Period)) * 100, 2) AS Pop_Growth_Rate,
    ROUND((Life_Expectancy - LAG(Life_Expectancy) OVER (PARTITION BY CSD ORDER BY Period)), 2)
    AS Life_Expectancy_Change
FROM (
    SELECT
        dem.CSD,
        dem.Period,
        SUM(dem.OriginalValue) AS Total_Population,
        AVG(le.OriginalValue) AS Life_Expectancy
    FROM demographics AS dem
    JOIN life_expectancy AS le
    ON dem.CSD = le.CSD AND dem.Period = le.Period
    GROUP BY dem.CSD, dem.Period
) AS combined_data
ORDER BY ABS(Pop_Growth_Rate) DESC, ABS(Life_Expectancy_Change) DESC
LIMIT 10;
"""
top_results = pd.read_sql_query(query_top_results, engine)
```

We calculated growth rates as the percentage change in population compared to the previous year. These rates were then paired with year-over-year changes in life expectancy to identify trends across municipalities. The results highlighted several important patterns.

Figure 5: Population Growth rate (x axis) compared to life expectancy (y axis) by municipality (colored bubbles) for the period between 2012 and 2023.

Figure 5 reveals that there isn't a significant pattern between the population growth rate and life expectancy changes. Most municipalities are clustered around the middle point from -20% to 20% population growth rate and -1 to + 1 years. Some insight from the outliers is that municipalities with much higher population growth often experienced stagnation or slight declines in life expectancy, likely due to strained local resources and healthcare systems. Urban centers like Calgary and Edmonton maintained higher life expectancy levels, aligning with their access to healthcare infrastructure. However, some smaller municipalities with modest growth demonstrated significant improvements, suggesting localized factors or interventions at play.

This exploration highlights the importance of sustainable growth strategies that align population changes with infrastructure and healthcare capacity. By adding a temporal dimension to the analysis, we provide a deeper understanding of the complex relationship between demographic shifts and life expectancy trends.

AGRICULTURE

The Agriculture industry, while largely forgotten about by urban dwellers, is certainly one of the most important industries to Albertans. With a GDP of 22.2 Billion dollars in 2020, Alberta accounted for largest contribution percentage of provinces to the Canadian national agriculture and livestock economy. While large and important, it is unclear how impactful the industry is on our life expectancy and air quality. To explore this topic, we decided to examine cattle, pigs, and cropland in the province.

Q. Is there a relationship between cropland, pigs, or cattle and our air quality?

Regarding air quality, it is a common assumption that the growth of plants and the raising of livestock have a limited impact on air quality. If anything, the growth of plants could improve air quality through photosynthesis. However, studies have demonstrated that, after the energy industry, the agriculture industry is the sector that produces the second most greenhouse gas emissions (St Pierre & McComb, 2022). To explore this apparent relationship between air quality-causing pollutants and the sector, we created four groups; Agricultural CSD’s (Municipalities with cropland), Non-Agricultural CSD’s (Municipalities without cropland), Pigs CSD’s (Municipalities with Pig livestock), and Cattle CSD’s (Municipalities with Cattle livestock). The Average Air Quality Index was identified for each of these four groups.

```
agri_aqi_1 = pd.read_sql_query('''
SELECT
    CASE
        WHEN (`Oats (Count)` > 0 OR `Canola (Count)` > 0 OR `Barley (Count)` > 0
        OR `Durum Wheat (Count)` > 0 OR `Non-Durum Wheat (Count)` > 0 OR `Other Crops (Count)` > 0)
        THEN 'Agricultural CSDs'
        ELSE 'Non-Agricultural CSDs'
    END AS CSD_Type,
    AVG(`Adjusted_AQHI`) AS Average_Adjusted_AQHI
FROM master_dataJF_2
GROUP BY CSD_Type;
''', engine)
agri_aqi_1
```

```
agri_aqi_2 = pd.read_sql_query('''
SELECT
    'Pigs CSDs' AS CSD_Type,
    AVG(`Adjusted_AQHI`) AS Average_Adjusted_AQHI
FROM master_dataJF_2
WHERE `Pigs (Count)` > 0;
''', engine)
agri_aqi_2
```

```
agri_aqi_3 = pd.read_sql_query('''
SELECT
    'Cattle CSDs' AS CSD_Type,
    AVG(`Adjusted_AQHI`) AS Average_Adjusted_AQHI
FROM master_dataJF_2
WHERE `Cattle and Calves (Count)` > 0;
''', engine)
agri_aqi_3
```

Our queries identified unique Average Air Quality values for each of our groups with Pigs CSDs having the lowest values at 2.09, Non-Agricultural CSDs having the highest values at 2.24, while Agricultural CSDs and Cattle CSDs had values between 2.21 and 2.22 respectively.

The largest outlier identified are the Pigs CSDs having a substantially lower value than the other groupings. Based solely on these values, the insight is that municipalities that have pig livestock have a better average air quality than municipalities without pig livestock. This insight is challenged, however, by the lack of a sufficient quantity of air quality data, and the ability to isolate different agricultural factors from the impact of other potentially influential factors.

Q. Do people who live in highly agricultural areas live higher quality lives?

Similar to the concept discussed in the question evaluating the relationship between agriculture and air quality, the groupings of municipalities will be leveraged to uncover if there is a relationship between quality of life and agriculture.

```
ag_LE_query_1 = pd.read_sql_query('''
SELECT
    CASE
        WHEN (`Oats (Count)` > 0 OR `Canola (Count)` > 0 OR `Barley (Count)` > 0
            OR `Durum Wheat (Count)` > 0 OR `Non-Durum Wheat (Count)` > 0 OR `Other Crops (Count)` > 0)
        THEN 'Agricultural CSDs'
        ELSE 'Non-Agricultural CSDs'
    END AS CSD_Type,
    AVG(`Life Expectancy (Years)`) AS Average_Life_Expectancy
FROM master_dataJF_2
GROUP BY CSD_Type;
''', engine)
ag_LE_query_1
```

```
ag_LE_query_2 = pd.read_sql_query('''
SELECT
    'Pigs CSDs' AS CSD_Type,
    AVG(`Life Expectancy (Years)`) AS Average_Life_Expectancy
FROM master_dataJF_2
WHERE `Pigs (Count)` > 0;
''', engine)
ag_LE_query_2
```

```
ag_LE_query_3 = pd.read_sql_query('''
SELECT
    'Cattle CSDs' AS CSD_Type,
    AVG(`Life Expectancy (Years)`) AS Average_Life_Expectancy
FROM master_dataJF_2
WHERE `Cattle and Calves (Count)` > 0;
''', engine)
ag_LE_query_3
```

Our queries identified unique Average Life Expectancy values for each of our groups. Similarly to our question regarding Air Quality, municipalities with Pig livestock are the outlying value with an average life expectancy of 79.71 years. Municipalities with Agriculture, No-Agriculture, and Cattle displayed average life expectancies of 79.89, 79.84, and 79.85 respectively. That municipalities with Pigs demonstrated the best average air quality and average life expectancy is worth questioning if there is a conjoining reason for this, or if this is merely a matter of coincidence.

Q. Has a relationship between Air Quality Index and Average Life Expectancy been identified for Municipalities with Pigs?

To effectively compare our results between Air Quality Index and Average Life Expectancy, we standardized our results for each using standard scaler from sklearn in python and displayed them in our figure below.

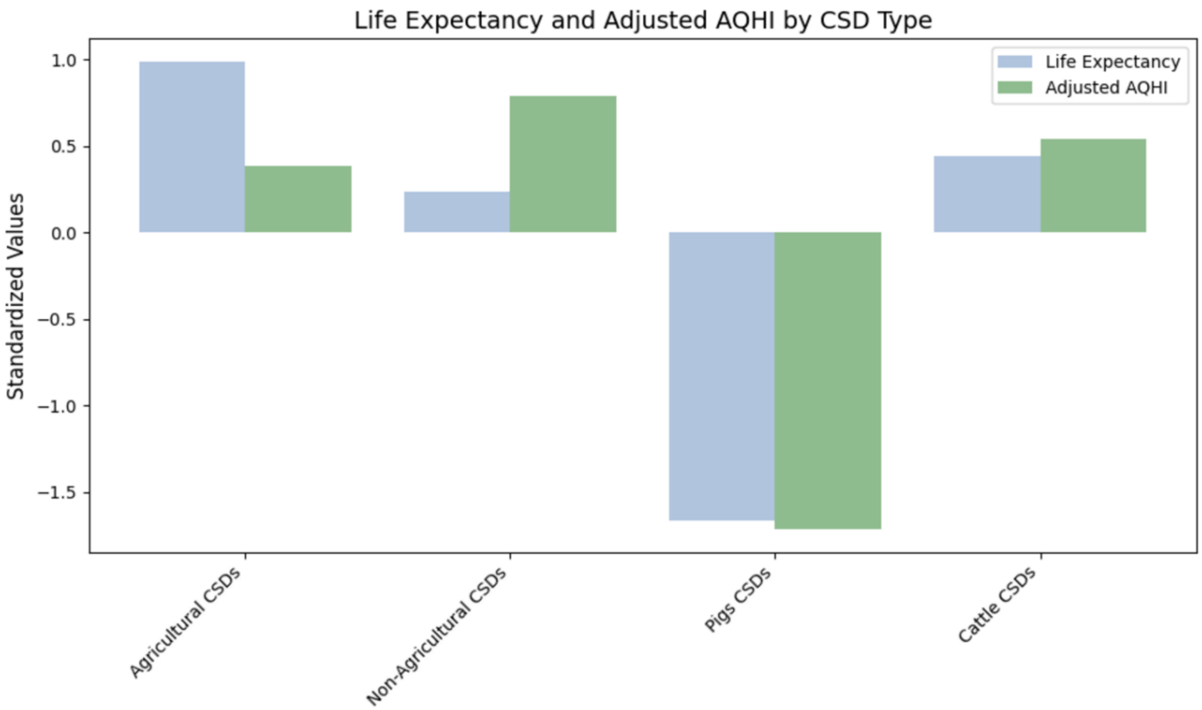

Figure 6: Life Expectancy and Adjusted AQHI by CSD Type

This figure demonstrates the interesting relationship between Life Expectancy and AQHI within our groupings of municipalities. Based solely on this plot, it appears that a relationship exists between the municipality groups and both of our dependent variables, especially for municipalities with Pigs. These municipalities have improved Air Quality, yet lessened life expectancy.

The statistical significance of this relationship is unclear. With so few AQHI data points and significant noise stemming from other potential predictor variables, it is improper to contribute the trends seen here as significant correlation. However, the trends between municipalities with Pigs and AQHI & Average Life Expectancy inspire the need for additional exploration with sufficient data.

AIR QUALITY INDEX & AVERAGE LIFE EXPECTANCY

The relationship between AHQI and average life expectancy was a pivotal aspect of our investigation. By analyzing the data, we aimed to identify whether regions with higher AHQI values experience lower life expectancy. By understanding this connection, we aimed to assess whether poor air quality significantly contributes to reduced life expectancy. This investigation provided insights into how environmental conditions, particularly air quality, intersect with public health outcomes and highlighted the broader implications for regional health disparities and policy-making.

Q. Is there any meaningful patterns that can be identified between air quality index and life expectancy?

To explore the relationship between the Air Quality Health Index (AQHI) and average life expectancy we started by plotting both variables.

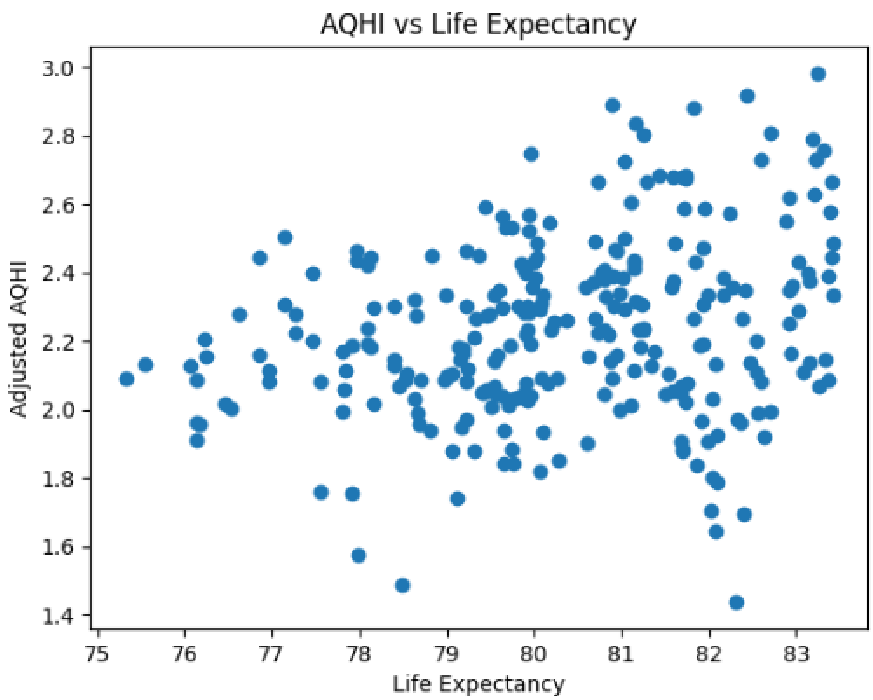

Figure 7: average life expectancy from 2012 to 2023 VS average AQHI for the same period.

The scatter plot reveals that some municipalities with the highest life expectancy (around 83 years old), also have the worst average AQHI (a reminder that 1 is the best and 10 is the worst AQHI). As explained earlier with the rest of our analysis the lack of valuable data suggests that the overall correlation is not strong enough to suggest a definitive causal relationship. This finding aligns with the understanding that AQHI, while significant, interreacts with other social environmental factors such as healthcare access, urbanization, and economic conditions influencing life expectancy. The analysis highlights the need for a multifaceted approach to understanding health outcomes, as AQHI alone may not fully explain variations in life expectancy.

```
query4 = pd.read_sql_query("""
    SELECT
       CASE
           WHEN Total_Population >= 100000 THEN 'Large (> 100,000)'
           WHEN Total_Population BETWEEN 10000 AND 99999 THEN 'Medium (10,000 - 100,000)'
           ELSE 'Small (< 10,000)'
       END AS population_group,
       AVG(`Life Expectancy (Years)`) AS mean_life_expectancy,
       AVG(Adjusted_AQHI) AS mean_aqi
    FROM masterdata
    GROUP BY population_group
    """, engine)

print(query4)
```

While we were unable to identify a linear relationship between air quality index and life expectancy, there were still some interesting correlations between the data that were found. Following the same reasoning as an earlier query, we wanted to explore the connection between life expectancy, air quality and population size. According to the government of Alberta, we consider a rural area to be a community outside of Calgary and Edmonton that has less than 100,000 residents (*Rural Renewal Stream – Eligibility*, 2024). All municipalities outside of these two cities (except for Lethbridge and Red Deer that have only surpassed the 100,000 threshold in the last few years) are consequently considered rural areas. Therefore, we once more grouped the municipalities and found some interesting results.

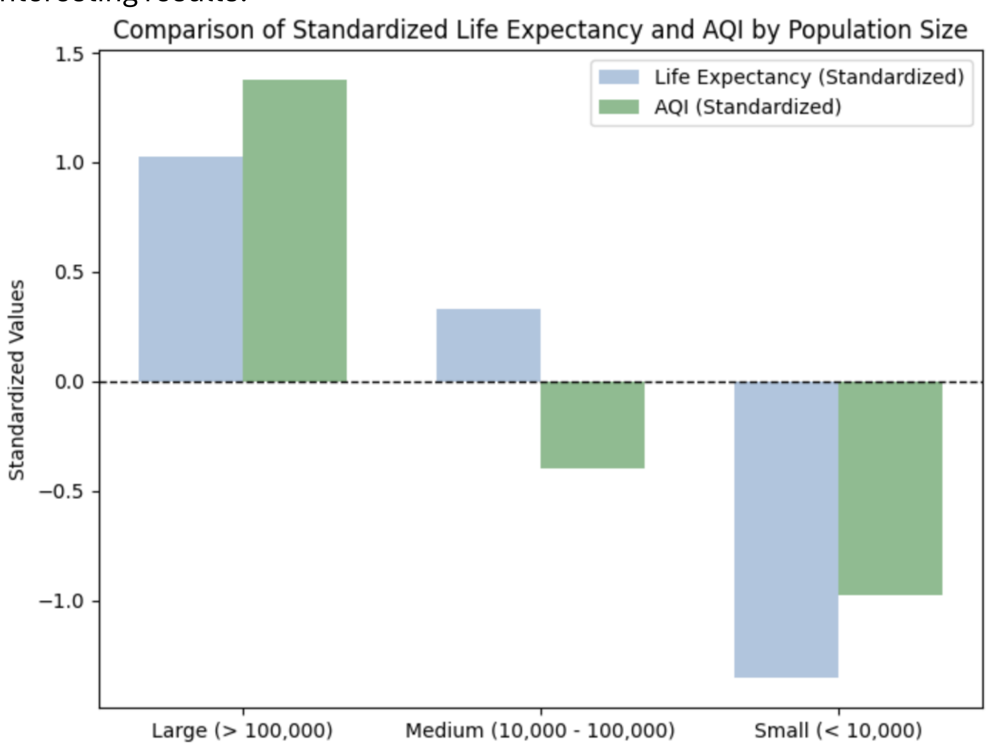

Figure 8: Side by side bar chart of standardized life expectancy and AQHI grouped by size of municipality

As mentioned earlier, we started by standardizing the AQHI and life expectancy from our query. When AQHI and life expectancy are on the same scale, we can see an interesting relationship across city sizes in Alberta. It appears that the largest cities have an average higher life expectancy, while simultaneously having a worse air quality. On the flip side, smaller municipalities have a significantly lower life expectancy but also have quite better air quality than big cities.

We found this interesting, and it was not a relationship we anticipated seeing. Intuitively, one would expect that worse air quality would either not affect life expectancy or potentially make life expectancy decline. Our correlation findings are the opposite. While Calgary and Edmonton showed an average life expectancy between 81 and 83 years in the last 12 years, rural areas were averaged between 79 – 80, and as the threshold for population size grew, the life expectancy started slowly catching up with Calgary and Edmonton. Previous studies on city demographics have shown that cities can be healthier places to live and have lower obesity rates than smaller cities (*City Living Makes You Healthier and Happier, According to Research*, 2024). Factors that may contribute to this include better access to medical care, increased walkability and transit opportunities, and more social opportunities. This relationship could be interesting to further explore in future studies.

CONCLUSION

In this project we were set to find the relationships between air quality health index and several environmental and demographic factors in Alberta with a strong support on multivariate regression analysis. After finding out first roadblock, we pivoted to life expectancy as the core of our project but soon after found that the statistical modelling would not be meaningful. With this in mind, we moved on to base our analysis on exploratory data analysis, visualization and most predominantly SQL queries. While modelling would’ve permitted us to make predictions, our current analysis only allows us to discuss current and past scenarios. Even with our EDA and queries, we couldn’t identify any definitive factors affecting AQHI or life expectancy. However, we did uncover some interesting insights and intersections between a variety of variables.

We discussed the alarming trends in wildfires and the impact of smoke for AQHI and we found out that life expectancy has been decreasing in Alberta since 2019. When looking at the energy sector, we could see that even though Alberta’s oil and gas industry is a significant contributor to greenhouse gas emissions, this doesn’t seem to have an impact on AQHI. We discussed that vehicle registration and residential construction don’t show a correlation with trends in AQHI. We focused on Calgary and Edmonton, and later on smaller municipalities, to review why urban centres have a higher life expectancy while simultaneously having a worse AQHI. And lastly, we addressed the fact that even though cattle and pigs are responsible for a significant portion of methane emissions that contributes to global warming, there was no sign of a correlation between this and life expectancy or AQHI.

While the end result wasn’t exactly what we were hoping for, this project taught us that unexpected results are not bad results. That data is data, and whether it provides what you are looking for or not, is not always within the user but in the integrity, quality and quantity of the data.

REFERENCES

About the air quality health index | alberta. Ca. (2024, November 28). https://www.alberta.ca/about-the-air-quality-health-index

Air Quality Health Index – Calculation. (2024, October 15). Alberta.ca. https://www.alberta.ca/air-quality-health-index-calculation

Air Quality Health Index – Calculation. (2024b, November 28). Alberta.ca. https://www.alberta.ca/air-quality-health-index-calculation#:~:text=AQHI%20in%20Alberta,-In%20Alberta%2C%20the&text=The%20following%20pollutants%20are%20considered,sulphide%20and%20total%20reduced%20sulphur

Alberta Regulatory Info. (n.d.). https://www.roadata.com/Regulatory/Alberta.aspx

Alberta's Oil Sands Resourceful. Responsible. (9 C.E.). In *https://open.alberta.ca/dataset/8c6d61b3-94e1-4940-b455-9451b05e507b/resource/3e25f3ae-016c-49de-8394-853bbee4174f/download/aenv-albertas-oil-sands-resourceful-responsible-7925.pdf*. Alberta Government.

BC Wildfire Service. (2024, September 16). *Wildfire season Summary - Province of British Columbia*. https://www2.gov.bc.ca/gov/content/safety/wildfire-status/about-bcws/wildfire-history/wildfire-season-summary

City living makes you healthier and happier, according to research. (2024, September 10). World Economic Forum. https://www.weforum.org/stories/2017/10/city-living makes-you-healthier-and-happier-study-finds/

Drew , A. (2023, June 27). What causes wildfires in Canada? [Journalism]. *The Narwhal*. https://thenarwhal.ca/canada-wildfires-cause/

Government of Canada, Canada Energy Regulator. (2023, November 24). *CER – results*. https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/2023/results/#:~:text=Description:%20This%20column%20chart%20shows,segments%20of%20the%20energy%20system.

Kopeliovich, J. (2023, June 9). *Climate Change, Canada's fires and US smoke - Woodwell Climate*. Woodwell Climate. Retrieved December 5, 2024, from https://www.woodwellclimate.org/climate-change-canadian-wildfires-smoke/?gad_source=1&gclid=Cj0KCQiAu8W6BhC-ARIsACEQoDBm0taqiIrDo2zyB5erzzAkj91FyFHu3u5fYdB9jKbwuutHaxU9k0QaApD7EALw_wcB

Ritchie, H., & Roser, M. (2024, March 18). Sector by sector: where do global greenhouse gas emissions come from? Our World in Data. https://ourworldindata.org/ghg-emissions-by-sector

Rural Renewal Stream – eligibility. (2024, November 28). Alberta.ca. https://www.alberta.ca/aaip-rural-renewal-stream-eligibility#:~:text=The%20AAIP%20defines%20a%20rural,communities%20and%20the%20designation%20process.

St Pierre, M., & McComb, M. (2022). Alberta has the highest farm operating revenues in Canada. In Canadian Agriculture at a Glance. Statistics Canada. Retrieved November 27, 2024, from https://publications.gc.ca/collections/collection_2022/statcan/96-325-x/CS96-325-2021-9-eng.pdf

Status of air quality in Alberta : air zones report - Open Government. (n.d.). https://open.alberta.ca/publications/alberta-air-zones-report

Wildfire smoke and your health. (n.d.). Government of Alberta. Retrieved December 5, 2024, from https://myhealth.alberta.ca/Alberta/Pages/wildfire-smoke-health.aspx

World Health Organization. (n.d.). Types of pollutants. Retrieved from https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants?utm

DATA SET SOURCES

Air quality index by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/air-quality-index-by-municipality

Building permits by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/building-permits-by-municipality

Cattle and calves by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/cattle-and-calves-by-municipality

Cropland by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/crop-acres-by-municipality

Dwelling units by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/dwelling-units-by-municipality

Greenhouse gas emissions by Municipality. Greenhouse gas emissions by municipality - Open Government. (2016, November 3). https://open.alberta.ca/opendata/greenhouse-gas-emissions-by-municipality

Historical wildfire data : 2006 to 2023 - Open government. (n.d.). [Dataset]. https://open.alberta.ca/opendata/wildfire-data

Housing starts by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/housing-starts-by-municipality#summary\

Life expectancy by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/life-expectancy-by-municipality

Major projects by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/major-projects-by-municipality

Motorized vehicle registrations by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/motorized-vehicle-registrations-by-municipality

Natural gas production by Municipality. Natural gas production by municipality - Open Government. (2016, November 3). https://open.alberta.ca/opendata/natural-gas-production-by-municipality

Pigs by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/pigs-by-municipality

Population by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/population-by-municipality

Residential share of property assessments by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/residential-share-of-property-assessments-by-municipality

Well count by municipality - Open Government. (n.d.). https://open.alberta.ca/opendata/well-count-by-municipality