STAA57 Group Project

What are the dominant research areas in Ontario's automotive sector, and how do institutions specialize in different fields?

Minh Tran - 1006804914, Rayaan Syed - 1010231081

March 15, 2025

I. Introduction

Research Goal

With the growing popularity of electric vehicles (EVs) and the rapid advancements in artificial intelligence (AI), particularly in the development of autonomous vehicles like Tesla, it is crucial for automotive research facilities in Ontario to focus their efforts and resources on the right areas. This will ensure a significant contribution to the academia world while assisting the government in implementing effective policies and educational institutions better educate, prepare the next generation of labors.

This report aims to examine the automotive research areas prioritizing by Ontario's research facilities and identify specialized facilities in trending fields. It will also explore which areas are attracting the most researchers and which institutions are specializing in particular research domains.

Dataset description

Data source

Our dataset was collected by researchers from teh Ministry of Economic Development, Job creation and Trade and was published on the Government of Ontario's public database in 2018.

Table 1: Data set Variables

```
## Rows: 545
## Columns: 37
## $ Institution                    <chr> "Brock University", "Carleton Univers~
## $ Researcher.Name                <chr> "Ahmed, Syed Ejaz", "Ahmadi, Mojtaba"~
## $ Associated.Facilities          <chr> "Centre for Statistical Consulting", ~
## $ Research.Areas                 <chr> "Traffic and road injury prevention."~
## $ Research.Chairs.Grant.Funding  <chr> "", "", "", "", "Canada Research Chai~
## $ Tag.1                          <chr> "Injury Prevention", "Mechatronics", ~
## $ Tag.2                          <chr> "", "Control", "Autonomy and AI", "Ne~
## $ Tag.3                          <chr> "", "", "Sensors", "", "", "", "", ""~
## $ Tag.4                          <chr> "", "", "", "", "", "", "", "", "", "~
## $ Tag.5                          <chr> "", "", "", "", "", "", "", "", "", "~
## $ Alternative.Fuels              <chr> "", "", "", "", "", "", "", "", "", "~
## $ Autonomy.and.AI                <chr> "", "", "x", "", "", "", "x", "", "",~
## $ Batteries.and.Fuel.Cells       <chr> "", "", "", "", "", "", "", "", "", "~
## $ Biocomposites                  <chr> "", "", "", "", "", "", "", "", "", "~
## $ Coatings.and.Corrosion         <chr> "", "", "", "", "", "", "", "", "", "~
## $ Connected.Vehicles             <chr> "", "", "", "", "", "", "", "", "", "~
## $ Control                        <chr> "", "x", "x", "", "", "", "", "", "",~
## $ Crashworthiness                <chr> "", "", "", "", "", "", "", "", "", "~
## $ Electronics                    <chr> "", "", "", "", "", "", "", "", "", "~
## $ Forming.and.Joining            <chr> "", "", "", "", "", "", "", "", "", "~
## $ High.Strength.Steel            <chr> "", "", "", "", "", "", "", "", "", "~
## $ Hybrid.and.Electric.Vehicles   <chr> "", "", "", "", "", "", "", "", "", "~
## $ Industrial.Processes           <chr> "", "", "", "", "", "", "", "", "", "~
## $ Injury.Prevention              <chr> "x", "", "", "", "", "", "", "", "", ~
## $ Internal.Combustion.Engines    <chr> "", "", "", "", "", "", "", "", "", "~
## $ Lightweight.Metals             <chr> "", "", "", "", "", "", "", "", "", "~
## $ Mechatronics                   <chr> "", "x", "", "", "", "", "", "", "", ~
## $ Nanotechnology                 <chr> "", "", "", "", "", "", "", "", "", "~
```

```
## $ Networks.and.Security            <chr> "", "", "", "x", "", "x", "", "x", "x~
## $ Noise..Vibration.and.Harshness   <chr> "Err:507", "Err:507", "Err:507", "Err~
## $ Polymers.and.Composite.Materials <chr> "", "", "", "", "", "", "", "", "", "~
## $ Powertrain                       <chr> "", "", "", "", "", "", "", "", "", "~
## $ Sensors                          <chr> "", "", "x", "", "x", "", "x", "", ""~
## $ Software                         <chr> "", "", "", "x", "", "", "", "", "", ~
## $ Stress.and.Fracture              <chr> "", "", "", "", "", "", "", "", "", "~
## $ Transportation.and.Charging      <chr> "", "", "", "", "", "", "", "", "", "~
## $ Vehicle.Design                   <chr> "", "", "", "", "", "", "", "", "", "~
```

Based on a summary of the data set, the `Noise..Viration.and.Harshness` variable was corrupted during the process of uploading or collecting the data and therefore will not be used in our report

```
# Remove variable with errors

researchers = researchers %>% select(!Noise..Vibration.and.Harshness )

# We removed Brock University, Lakehead University, Royal Military College, and Laurentian University f

researchers <- researchers %>% filter(!Institution %in% c("Brock University", "Lakehead University", "La
```

Some of the main categorical variables that can help categorizing the data include

- **Researcher.Name**
- **Institution**: Name of university or research center that the researcher works at
- **Research.Areas**: Researcher general field of research
- **Tag**: Researcher's specialized field of research

In addition, the data set also contains research fields as variable (e.g. Alternative Fuels, Autonomy and AI, Vehicle Design) to indicate whether or not a researcher work fall into that field of research

II. Data Overview Analysis

Descriptive Statistic

From our dataset, we are able to extract the disparity of researchers across fields of research. This would include research fields such as "Network and Security", "Autonomy and AI", "Nanotechnology", and other vehicle-related fields.

```
## # A tibble: 13 x 29
##    Institution                Networks and securit~1 `Autonomy and AI` Software
##    <chr>                                       <int>             <int>    <int>
##  1 "Carleton University"                          37                23       21
##  2 "University of Waterloo"                       22                16        7
##  3 "McMaster University"                           1                 4        7
##  4 "University of Toronto"                         6                11        1
##  5 "Ryerson University"                            6                 6        2
##  6 "University of Windsor"                         5                 5        0
##  7 "Queen's University"                            2                 1        2
```

```
##  8 "University of Guelph"                          0                4        0
##  9 "University of Ontario Ins~                     7                2        6
## 10 "University of Ottawa"                          4                4        1
## 11 "York University"                               0                2        0
## 12 "University of Western Ont~                     2                0        0
## 13 "Ryerson University   "                         0                0        0
## # i abbreviated name: 1: `Networks and security`
## # i 25 more variables: `Connected vehicles` <int>, Nanotechnology <int>,
## #   `Polymers and composite materials` <int>,
## #   `Hybrid and electric vehicles` <int>, Powertrain <int>,
## #   `Batteries and fuel cells` <int>, Sensors <int>,
## #   `Transportation and charging` <int>, `Forming and joining` <int>,
## #   `Injury Prevention` <int>, Electronics <int>, ...
```

Upon reviewing the table, the three leading researching fields are "Networks and Security", "Autonomy and AI", and "Polymers and Composite Materials". The universities have reported 50+ researchers working in these ares which would indicate there is significant progress being made. On the contrast, "Crashworthiness" and "Mechatronics" are among the least studied fields with less than 10 researchers total.

We now check for the dominating research field for each university. The table below is generated by finding the max researchers in an institution across a research field.

```
## # A tibble: 12 x 2
## # Groups:   Institution [12]
##    Institution                                          Research_Area
##    <chr>                                                <chr>
##  1 Carleton University                                  Networks and security
##  2 Mcmaster University                                  Powertrain
##  3 Queen's University                                   Lightweight metals
##  4 Ryerson University                                   Transportation and char~
##  5 University Of Guelph                                 Batteries and fuel cells
##  6 University Of Ontario Institute Of Technology (Uoit) Networks and security
##  7 University Of Ottawa                                 Sensors
##  8 University Of Toronto                                Polymers and composite ~
##  9 University Of Waterloo                               Networks and security
## 10 University Of Western Ontario                        Forming and joining
## 11 University Of Windsor                                Lightweight metals
## 12 York University                                      Transportation and char~
```

From the table, the universities are associated with their highest producing research field. Several universities are mostly researching "Networks and Security" like Carleton University, University of Waterloo, University of Ontario Institute of Technology.

This information can also be crucial to understand how the fields with low researchers from the previous table are not being targeted enough by the universities.

We now check for the percentage breakdown of the university's dominating research field and their contribution as a whole. This is important to understand the concentration of university's in specific fields and whether some diversification can be made to the university.

```
## # A tibble: 12 x 4
## # Groups:   Institution [12]
##    Institution                                 Research_Area Count Percentage
##    <chr>                                       <chr>         <int>      <dbl>
##  1 Carleton University                         Networks and~    37      24.5
```

```
##  2 York University                                  Transportati~     6      22.2
##  3 University Of Ottawa                             Sensors           7      21.2
##  4 University Of Western Ontario                    Forming and ~     3      17.6
##  5 Ryerson University                               Transportati~    11      17.5
##  6 Queen's University                               Lightweight ~     7      15.2
##  7 University Of Guelph                             Batteries an~     7      12.5
##  8 Mcmaster University                              Powertrain       16      11.0
##  9 University Of Toronto                            Polymers and~    13       9.92
## 10 University Of Windsor                            Lightweight ~    11       9.65
## 11 University Of Ontario Institute Of Technology~ Networks and~     7       9.21
## 12 University Of Waterloo                           Networks and~    22       7.80
```

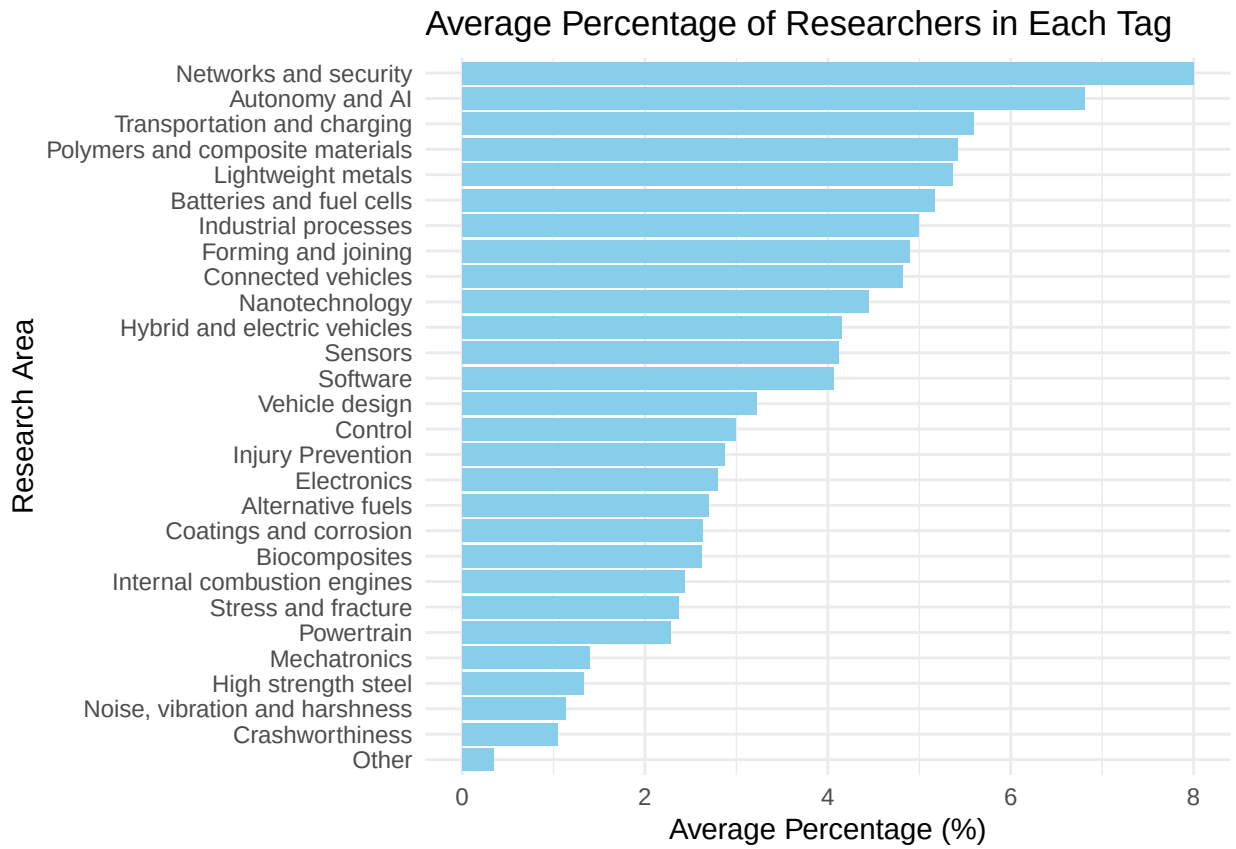
From the table, we see that Carleton University, York University, and University of Ottawa have 20%+ of their research contributions among their highest fields. Waterloo University and University Of Ontario Institute Of Technology have the most diversification as well as high amounts of researchers across several fields.

Graph & Visualizations

- Bar chart: Research distribution across institution
- Pie chart: Proportion of of different research fields
- History, frequency of research tag

III. Statistical Analysis

Let's pick our highest leading tag and see how it performs in bootstrapped data. We want to observe how it does compared to other tags and if it's the leading tag in all the samples by takign an average.



In this graph, the data of the institutions and their tags have been replicated 1000 times. We take the average percentage to see how they perform over the 1000 trials and gain a graph of each tags performance.

```
# Confidence Interval for "Networks and security"
network_security_percentage = boot_results_df %>%
  filter(Research_Area == "Networks and security") %>% pull(Percentage)

# 95% confidence interval
lower_ci = quantile(network_security_percentage, 0.025)
upper_ci = quantile(network_security_percentage, 0.975)

cat("95% Confidence Interval for Network and Security: [", lower_ci, ", ", upper_ci, "]\n")
```

```
## 95% Confidence Interval for Network and Security: [ 6.468531 ,  9.702797 ]
```

By taking a confidence interval for Network and Security, we observe that the true percentage of researchers in Network and Security lies in the interval [6.468531, 9.702797]. This is a very high value when comparing this to the other tags in our horizontal bar chart.

Let's examine how Networks and Security performs when compared to the 2nd highest tag (Autonomy and AI) in our bootstrapped data to see if it always leads the data by a hypothesis test.

```
# 2nd-ranked tag
second_place_tag = average_percentage_df$Research_Area[2]

# Prepare data for paired comparison
```

```r
paired_data = boot_results_df %>%
  filter(Research_Area %in% c("Networks and security", second_place_tag)) %>%
  pivot_wider(names_from = Research_Area, values_from = Percentage)

# Paired t-test (one-sided: Networks and Security > 2nd place)
t_test_result = t.test(
  paired_data$`Networks and security`,
  paired_data[[second_place_tag]],
  paired = TRUE,
  alternative = "greater"  # Tests if Networks and Security > 2nd place
)

cat("Paired t-test: Networks and Security vs.", second_place_tag, "\n")
```

```
## Paired t-test: Networks and Security vs. Autonomy and AI
```

```r
cat("Mean % Network and Security:", mean(paired_data$`Networks and security`), "\n")
```

```
## Mean % Network and Security: 8.001923
```

```r
cat("Mean % (", second_place_tag, "):", mean(paired_data[[second_place_tag]]), "\n")
```

```
## Mean % ( Autonomy and AI ): 6.806818
```

```r
cat("T-statistic:", t_test_result$statistic, "\n")
```

```
## T-statistic: 31.29503
```

```r
cat("P-value:", t_test_result$p.value, "\n")
```

```
## P-value: 1.070572e-150
```

```r
if (t_test_result$p.value < 0.05) {
  cat("Reject H . Networks and Security is significantly higher than the other tags. (p < 0.05).")
} else {
  cat("No significant difference (p   0.05).")
}
```

```
## Reject H . Networks and Security is significantly higher than the other tags. (p < 0.05).
```

Based off our paired t-test, we get a average Network Security score of 8.001923 and a average Autonomy and AI score of 6.806818. Since the averages are distant, we can expect a large t statistic and a small p value.

We observe a t statistic value of 31.29503 and a extremely small p value 1.070572e-150 ($p < 0.05$), rejecting the null hypothesis and observing Networks and Security is significantly higher than the other tags.

IV. Predictive modeling (Regression)

Build a regression model to predict number of researcher in a field, institutions

Build a regression model to see if an institution is likely to have to a grant funding

Using cross-validation to validate the 2 models built above

V. Summary

Appendix