**Research Question**:

Classification approach: which neighborhoods tend to have different types of calls and such

Time series approach: trying to predict number of calls per day

**Response for classification**:

Probability of type of call (medical, fire, ...)

Probability of neighborhood

**When**: 2023

**Where**: Evansville

**Data Sources:**

Weather: <https://www.wunderground.com/history/daily/us/in/evansville/KEVV/date/2023-7-1>

Drive time: <https://chandoo.org/wp/distance-between-places-excel-maps-api/>

**Variables**: Station that took call, call reason, time to respond/arrive, date, address (distance to station, area: neighborhood), number of people per station, number of trucks per station, time to completion, reignition/return to same address, multiple stations respond to same address, allocation of positions/jobs per run, weather data, traffic data

**Other thoughts:**

What if two calls occur near the same time or at the same time?

Want to use LaTex or Quarto to write paper?

Get Quarto: <https://quarto.org/docs/get-started/>

Quarto R Tutorial: <https://quarto.org/docs/get-started/hello/rstudio.html>

Quarto Writing math/equations: <https://quarto.org/docs/visual-editor/technical.html>

Evansville FD:

[**https://www.evansvillegov.org/city/topic/index.php?topicid=976&structureid=20**](https://www.evansvillegov.org/city/topic/index.php?topicid=976&structureid=20)

EFD Still Territories: [**https://www.google.com/maps/d/u/0/viewer?mid=1cz02\_fizDNde-Ar-1Yq5lQk8tyw&hl=en\_US&ll=37.9736183555321%2C-87.54371350000001&z=12**](https://www.google.com/maps/d/u/0/viewer?mid=1cz02_fizDNde-Ar-1Yq5lQk8tyw&hl=en_US&ll=37.9736183555321%2C-87.54371350000001&z=12)

**Methods**:

**Joey: classification**

* Decision trees
  + Bagging
  + Random forest
  + Conditional random forest
  + Bayesian regression and classification trees
  + Boosting
* Logisitic or multinomial regression
* Kernal estimation
* SVM (support vector machine)
* Fisher’s Discriminant Analysis
* Naive Bayes
* Neural Networks
* KNN ( k nearest neighbors)

**Ryan: time series**

* Might be able to use regression, but can check Durbin-Watson statistic (independence assumption) and check for multicollinearity: [https://medium.com/@kirudang/multivariable-time-series-approach-guide-for-time-series-with-multiple-predictors-ac89c5893370](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fmedium.com%2F%40kirudang%2Fmultivariable-time-series-approach-guide-for-time-series-with-multiple-predictors-ac89c5893370&data=05%7C02%7Chlcook1%40usi.edu%7Cd53a2645899f423b809a08dc18970c4b%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638412284216256267%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=2gNg%2Bxqh07u9Al5Ex8Zc6M6TLtiHQK4kxyDkDRx5148%3D&reserved=0)
* Could use XGBoost: <https://www.kaggle.com/code/robikscube/tutorial-time-series-forecasting-with-xgboost/notebook>
* <https://community.rstudio.com/t/best-option-to-predict-time-series-with-multiple-variables/49949/11>
* Multivariate Time Series: [https://medium.com/analytics-vidhya/a-multivariate-time-series-guide-to-forecasting-and-modeling-with-python-codes-8733b5fd1a56](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fmedium.com%2Fanalytics-vidhya%2Fa-multivariate-time-series-guide-to-forecasting-and-modeling-with-python-codes-8733b5fd1a56&data=05%7C02%7Chlcook1%40usi.edu%7C2606fc2925034ee9c1c708dc18975f45%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638412285606700654%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=G83omizQSdV7Q4nX0oCutyk1M%2FQkKGW8mjlbRDfmXxk%3D&reserved=0)
* VAR model from Penn state: <https://online.stat.psu.edu/stat510/lesson/11/11.2>
* (ARMAX and VARMA) Times Series and Its Applications Book: <http://www.stat.ucla.edu/~frederic/415/S23/tsa4.pdf>

Questions to ask Tony Knight:

1. What do the abbreviations in nature mean?
2. What are prime unit (main unit that went out? Station that responded to the original call?) and employee (captain of the unit that originally responded?)? Are they related to each other?
3. Is there any other data that can be attached to the incidents, such as time to respond, time to complete, etc.

**Other Data Sources:**

Data World Datasets: <https://data.world/datasets/fire-department>

IN Government Data Dashboard (data could be manually collected): <https://www.arcgis.com/apps/dashboards/3bf2ccce1aa04597a5b1c70d50ba137a>

(comes from <https://www.in.gov/dhs/data-and-dashboards/data-and-statistics/>)

How to request data (although may take too long): <https://www.in.gov/mph/request-data/#Row_level_De_identified_Data>

EMS runs 2014-2022: <https://hub.mph.in.gov/dataset/emergency-medical-service-runs>

Can filter before download (2022 has over 1 million rows for all of IN)

**January 12th Notes:**

Have per county in IN high quality capacity, iread scores per county in IN (but this is given to 3 graders, start of 3rd grade)

**Potential Questions to Answer:**

**Sources:**

1. Early Learning Indiana’s Closing the Gap

Url: https://earlylearningin.org/closing-the-gap/#interactive-map

This has high-level county profiles.

2. Brighter Futures Indiana Data Center

Url: https://brighterfuturesindiana.org/data-center

3. IU’s data hub

Url: https://www.stats.indiana.edu/index.asp

There is a lot of US census info here broken down by counties.

4. Grow Database

Url: https://savi.growcapacity.info/

Provides demographics.

5. can request data from IN although unsure on time frame

https://www.in.gov/doe/legal/data-requests/

6. **Childcare inspections.** “Across 41 states, one in ten licensed daycare facilities is overdue for an inspection,” according to an [analysis by USAFacts](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fusafacts.org%2Fdata-projects%2Fchildcare-inspections&data=05%7C01%7Chlcook1%40usi.edu%7C79ca3a2067c1437a574308dbdada7ea2%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638344404278311838%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=60FEA38q%2B5ZnRFenyt7pUtbeXydMaxiAfPv%2F9wb0vsM%3D&reserved=0), which gathered inspection-timing data for 148,000 facilities. The analysis’s downloadable datasets indicate the number of facilities overdue (versus on-time or unknown) in each state and county, as well as the distribution of days-since-last-inspection in each state. Most of the inspection data had to be programmatically collected from state portals, [according to Amber Thomas](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.linkedin.com%2Fposts%2Famberrthomas_are-states-meeting-their-childcare-inspection-activity-7120861735833518080-V-6s%2F&data=05%7C01%7Chlcook1%40usi.edu%7C79ca3a2067c1437a574308dbdada7ea2%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638344404278311838%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=iYbKvLybFAsodXGSb%2BaIPsacvdpmfO%2BY3P7m2a0slLI%3D&reserved=0), except for California, which [publishes bulk data](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ccld.dss.ca.gov%2Fcarefacilitysearch%2FDownloadData&data=05%7C01%7Chlcook1%40usi.edu%7C79ca3a2067c1437a574308dbdada7ea2%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638344404278311838%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=2UT419hODyBBb4MdsILmOSMQqAsBlQLs0rkuaVxk6qc%3D&reserved=0) (including inspection dates) about all childcare facilities.

7. **Childcare prices.** The [National Database of Childcare Prices](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.dol.gov%2Fagencies%2Fwb%2Ftopics%2Ffeatured-childcare&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389808099%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=S3%2B8WyR4fG84pMd1MnHmM1vmDlxFITJAUcU7Yz3MRcg%3D&reserved=0), launched [in January](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.dol.gov%2Fnewsroom%2Freleases%2Fwb%2Fwb20230124&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389964328%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=34KY9eUnLSGmFd3k%2FBCqf1dhplSWFz1SNqvhyqFDbGw%3D&reserved=0) by the Department of Labor’s [Women’s Bureau](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.dol.gov%2Fagencies%2Fwb&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389964328%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=U%2BJ8jx72ewdAR0bTxCglV1OPIH9QKU65JkKEE4GNXKI%3D&reserved=0), “is the most comprehensive federal source of childcare prices at the county level.” For each county and year from 2008 to 2018, the dataset provides estimates of the median and 75th-percentile weekly cost, disaggregated by provider type and child age. The estimates are [calculated from the market surveys](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.dol.gov%2Fsites%2Fdolgov%2Ffiles%2FWB%2Fmedia%2FNationalDatabaseofChildcarePricesTechnicalGuideFinal.pdf&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389964328%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=nqWJA3DJLNLYiRBqS%2B0NnO6eOvLWt%2BNm9YRpltKxxZA%3D&reserved=0) the federal Child Care and Development Fund [requires participating states to conduct](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fchildcareta.acf.hhs.gov%2Fequal-access-and-market-rate-surveys&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389964328%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=1iDaa2oeXQIyNDqE7G2S8lqXBLwEqDRMBhBCxmNEMfI%3D&reserved=0). [h/t [Erik Gahner Larsen](https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgithub.com%2Ferikgahner%2FPolData%2Fcommit%2F062d1fc66219f9f954ed66592d2633477a40c264&data=05%7C01%7Chlcook1%40usi.edu%7C23afd23decda480f990208db46501fea%7Cae1d882c786b492c90953d81d0a2f615%7C0%7C0%7C638181082389964328%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=uB5yys37dUhSwgwDMYH0YmMTPnK9v3NnpfBRVXXdcGQ%3D&reserved=0)]

8. Kindergarten readiness study

Url: <https://www.in.gov/fssa/carefinder/files/OMWPK-2021-Kindergarten-Readiness-Final-Report-0122.pdf>

9. Kindergarten readiness info from NORC

<https://www.norc.org/research/projects/prek-on-my-way-evaluation.html>

<https://www.norc.org/research/projects/kindergarten-readiness-indicators.html>

<https://www.the74million.org/article/closing-the-data-gap-for-indianas-littlest-learners-a-model-for-other-states/>

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.in.gov/fssa/carefinder/files/KRI2022AssessmentReport.pdf

2023 Spring-Summer Combined IREAD 3 Results

<https://www.in.gov/doe/files/iread3-final-corporation-and-school-results-2023.xlsx>

**January 8th and 10th Notes:**

Joey to ask Building Blocks about access to data and if they have questions they want answered

28 counties, license capacity (how many kids caregivers can watch), how many quality seats, type of childcare (ministries, licensed, ...), costs, retention of workers in centers

Goal: predicting high quality capacity (% children in high quality care)

Ryan look up COVID influences on ACT/SAT scores, drop out rates, college enrollment

could ask for USI for their data on acceptance/applications

Topics: time series, finance/housing, trends, predicting events (regression, classification), second language acquisition

Trends: resistant to drastic changes (floral patterns), good investments, also what makes something trendy (profitable)

Real data, big data

NOT SPORTS

Ryan’s past courses: regression, categorical data analysis, design and analysis of experiments, math stats, prob, time series, sampling

Joey’s past courses: regression, design and analysis of experiments, math stats, prob, time series, sampling, data mining, multivariate