Volunteering with DataKind: Predicting Foster Care Case Complexity

Kevin Kho (kdykho@gmail.com)
Prepared for Orlando Machine Learning and Data Science

About Me

- Kevin Kho
- BS/MS Civil Engineering from University of Illinois at Urbana-Champaign
- 1 year research with USGS
- 2 years as a data scientist with Itron
- 1 year as a data science mentor with Thinkful
- How I ended up volunteering

Acknowledgements

- DataKind
- CBCCF - Matt Baker
- Microsoft Cities
- Volunteer Team:
  - Anasuya Das, Alice Feng, Margaret Furr, Juan Bages, Sebastian Ouellet, Matt Robinson, Sebastien Ouellet, Sajala Shukla, Min Li
- ** Most of the slides for this presentation were the work of Alice Feng

Community Based Care of Central Florida

- Non-profit organization responsible for foster care, adoption, and diversion
- Since 2004, CBCCF has facilitated 1000 adoptions, reduced child removal rate by 25%, successfully reunited 69% of families, and doubled the number of foster homes
- Also have programs for minors to get livelihood training

In 2016

JOB STRESS

Caseload Overload

Job Trauma

Low Pay

Meant

1 IN 7 LEFT THEIR JOB

Costing

$54K TO REPLACE A SINGLE WORKER

Causing

KIDS to WAIT

- Child's case must be re-assigned.
- New caseworker must have 102 hours of core training prior to carrying full caseload.
- Must then re-establish child & family relationships.

. . . meanwhile, the child waits.

CASEWORKER TURNOVER IS ONE OF THE BIGGEST CONTRIBUTORS TO KIDS GETTING STUCK IN FOSTER CARE.

Community Based Care of Central Florida

- In a case study conducted by Milwaukee County, children who have one caseworker achieve permanency 74.5 percent of the time. With two caseworkers the chance of permanency decreases to 17.5 percent
- Community Based Care of Central Florida, where training new caseworkers costs between 25%-40% of their annual salary

Community Based Care of Central Florida

- Met with DataKind to scope projects to increase retention of caseworkers
- 2 Projects came out:
  a. Planning Optimization Tool
  b. Measurement of Case Complexity

Expected Timeline

October 2017	Team Formation
Early December 2017	Project Start
Late January 2018	EDA
March 2018	Initial Deliverables
May 2018	Final Deliverables
June 2018	Handoff

Actual Timeline

October 2017	Team Formation
Early December 2017	Project Start
February 2018	EDA
April 2018	Initial Deliverables
July 2018	Final Deliverables
August 2018 (2 months)	Handoff

Project Structure

- Project Manager
- Data Ambassador
- 2 Tech Leads
  - 3 Members Each

Relationship of Two Projects

- Geospatial Component
- Cases with siblings
- Where does the caseworker already travel to?

Scheduling Optimization Tool

- Case workers are the intended user
- The tool helps case workers decide how to sequence tasks
- Users log into the tool once per week
- **Solutions should be flexible**
- The system may be able to leverage information about activity on a particular case to identify upcoming due dates

Scheduling Optimization Tool

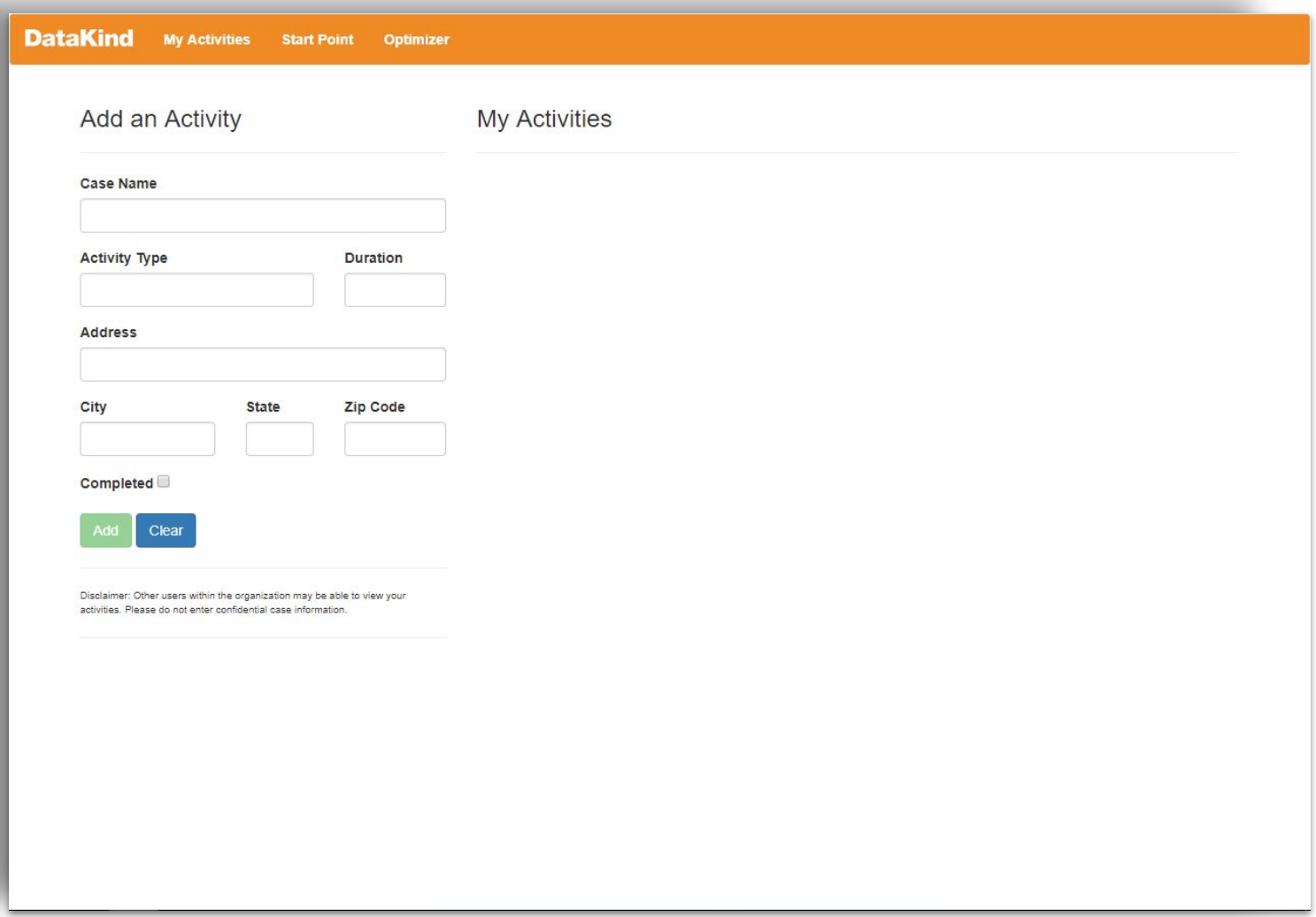
- Angular
  - Computes an optimal schedule for a case worker, given the case worker's activities and constraints
  - Allows users to modify a schedule after it has been optimized
  - Pulls in individual user information based on an id in the URL
  - Modified baseline Angular configurations to ensure that it is compatible with **Internet Explorer**

Scheduling Optimization Tool

- Built a backend application using Python and MySQL
  - Delivered a unit tested Python package that includes web services, and interface with the *COIN-OR* integer programming solver, and utilities to help manage the application
  - MySQL database with an automated build process

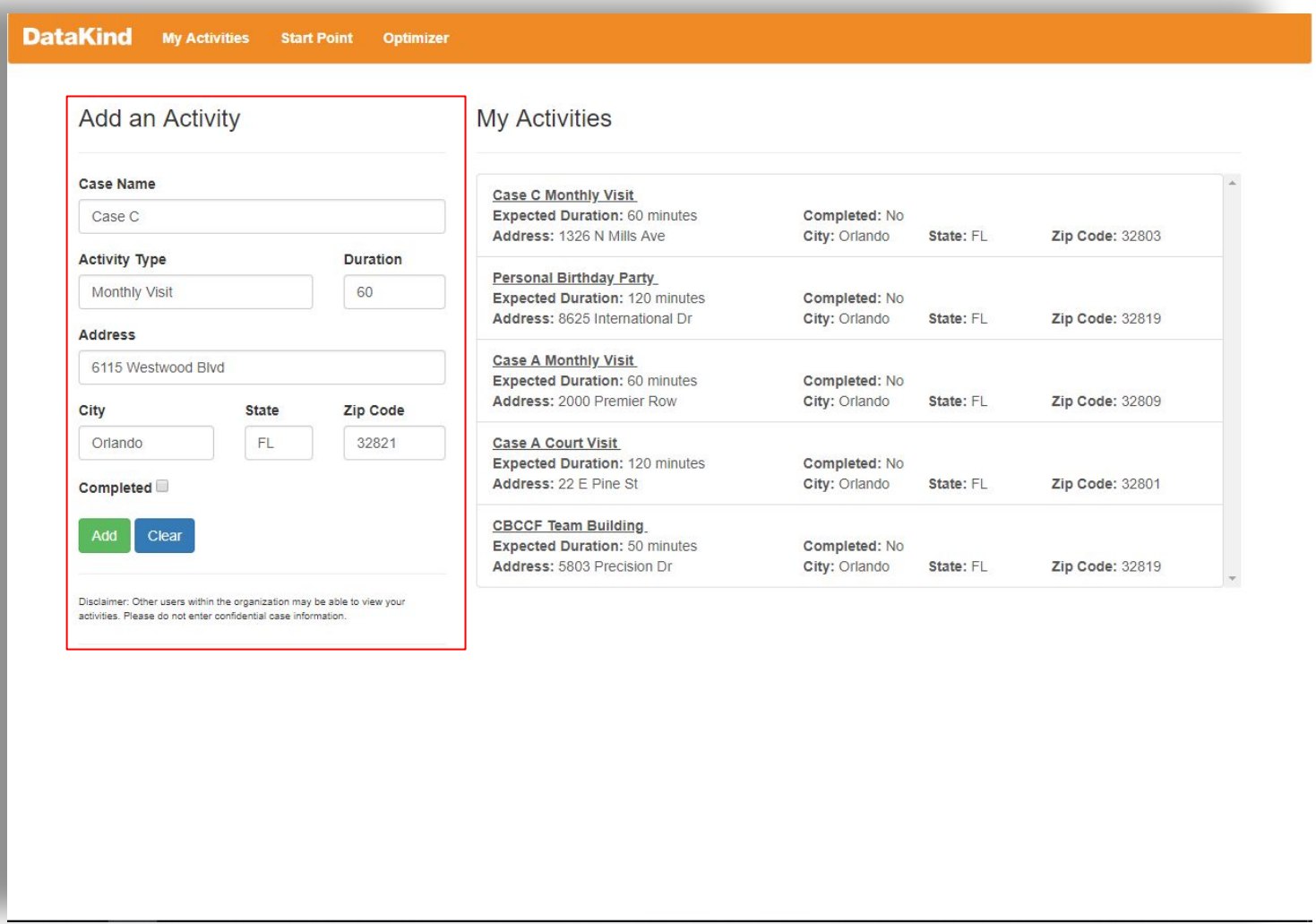
Optimization App Features

User Interface



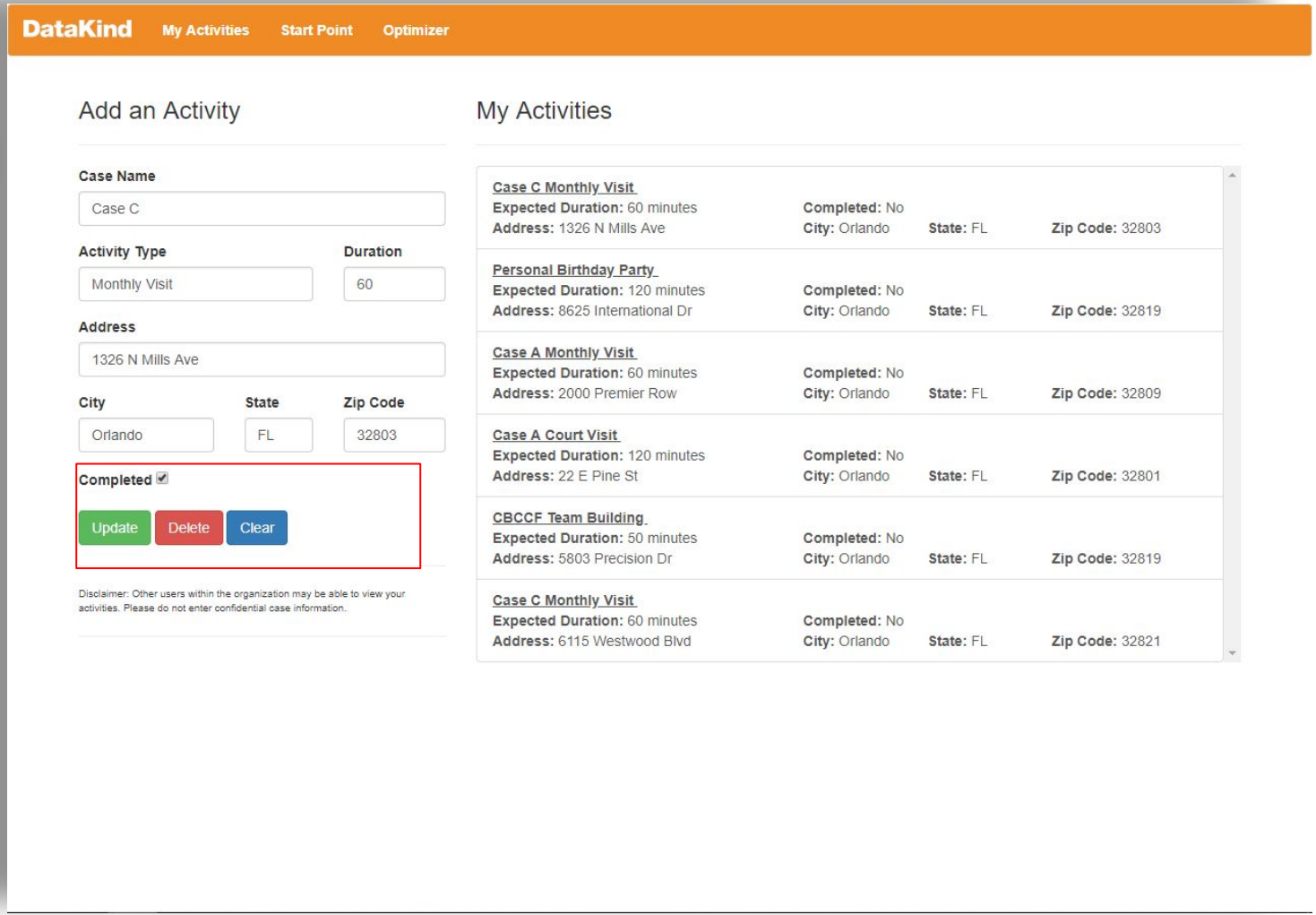
DataKind

Adding Activities



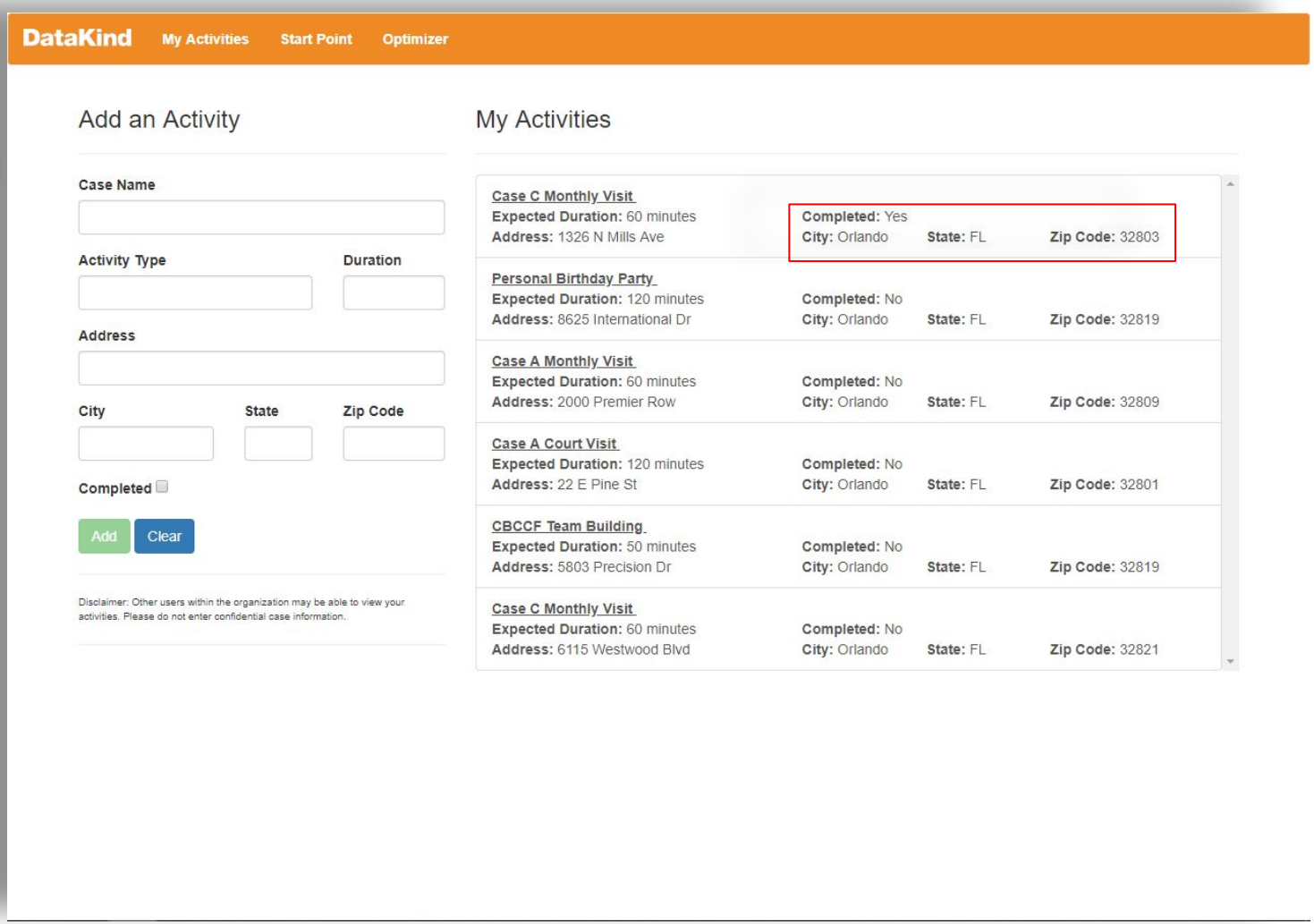
DataKind

Marking Completed Activities



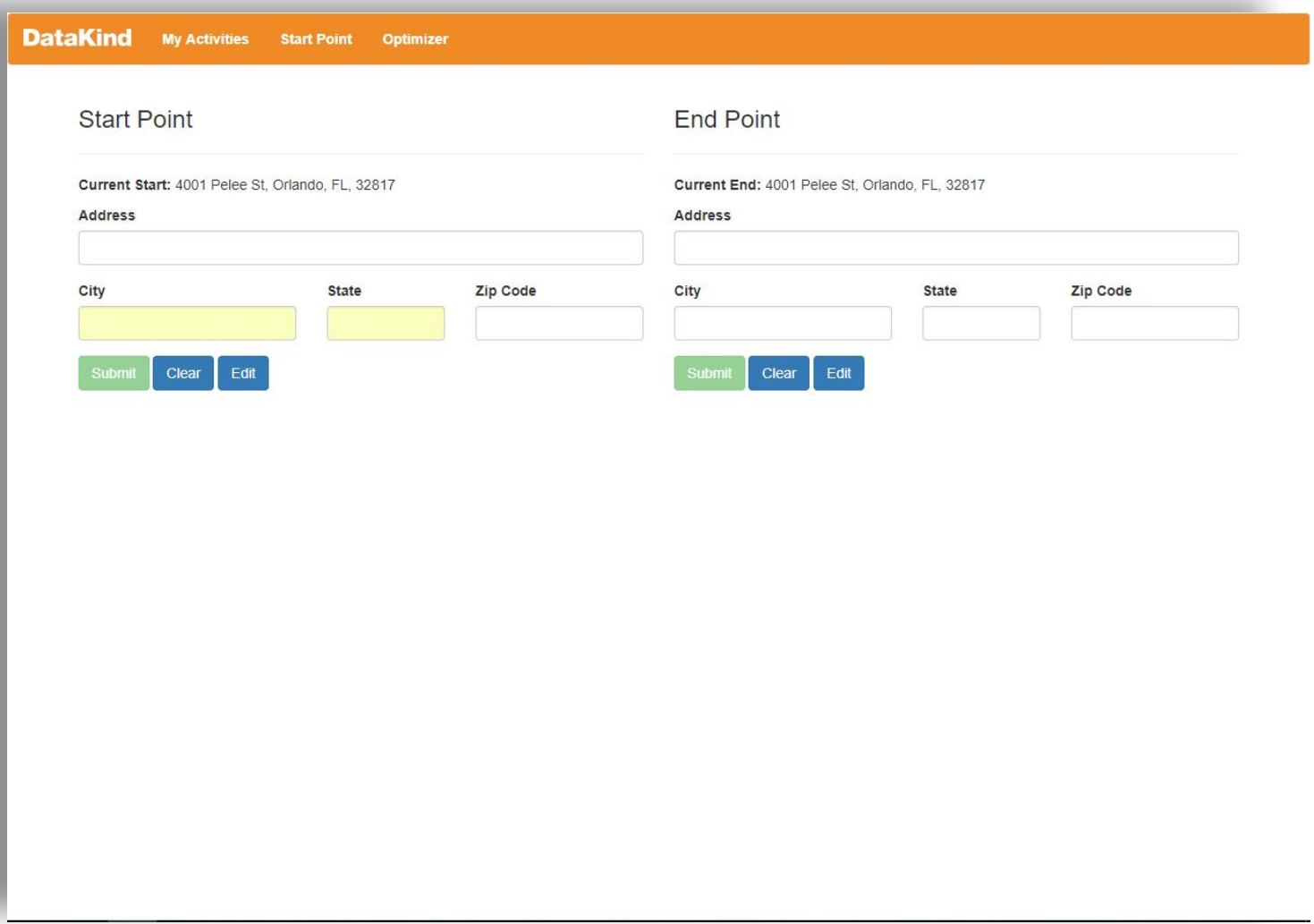
DataKind

Marking Completed Activities



DataKind

Setting Starting/ Ending Points

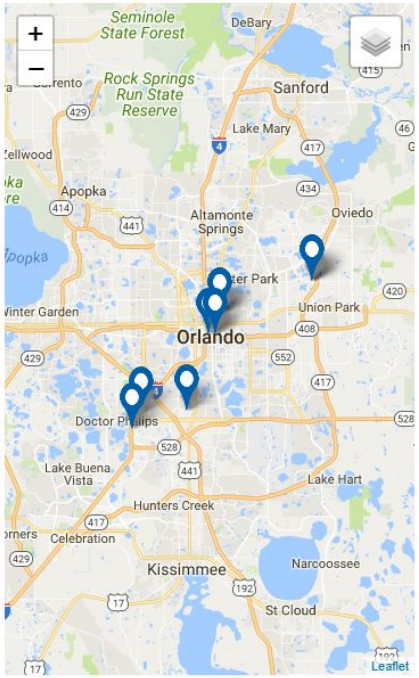


DataKind

Optimizer

DataKind | My Activities | Start Point | Optimizer

Map



Schedule Optimizer

Schedule | Daily Hours | Assignments

Optimize | myaddr@example.com | E-mail

Monday

Time: 4.8 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
Case A Monthly Visit	2000 Premier Row, Orlando, FL, 32809	60 mins	21 mins	⌄
Personal Birthday Party	8625 International Dr, Orlando, FL, 32819	120 mins	10 mins	⌃ ⌄
CBCCF Team Building	5803 Precision Dr, Orlando, FL, 32819	50 mins	5 mins	⌃ ⌄
Finish	4001 Pelee St, Orlando, FL, 32817	-	20 mins	-

Tuesday

Time: 5.6 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
Case A Court Visit	22 E Pine St, Orlando, FL, 32801	120 mins	13 mins	⌃ ⌄

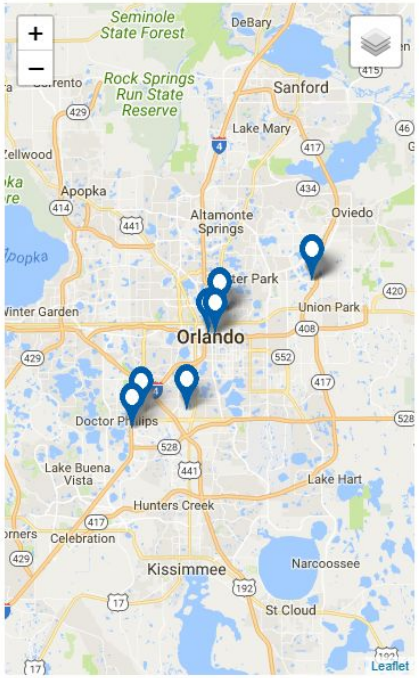
DataKind

Ordering Activities

DataKind

DataKind | My Activities | Start Point | Optimizer

Map



Schedule Optimizer

Schedule | Daily Hours | Assignments

Optimize | myaddr@example.com | E-mail

Monday

Time: 4.9 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
CBCCF Team Building	5803 Precision Dr, Orlando, FL, 32819	50 mins	20 mins	⌄
Case A Monthly Visit	2000 Premier Row, Orlando, FL, 32809	60 mins	8 mins	⌃ ⌄
Personal Birthday Party	8625 International Dr, Orlando, FL, 32819	120 mins	10 mins	⌃ ⌄
Finish	4001 Pelee St, Orlando, FL, 32817	-	23 mins	-

Tuesday

Time: 5.6 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
Case A Court Visit	22 E Pine St, Orlando, FL, 32801	120 mins	13 mins	⌃ ⌄

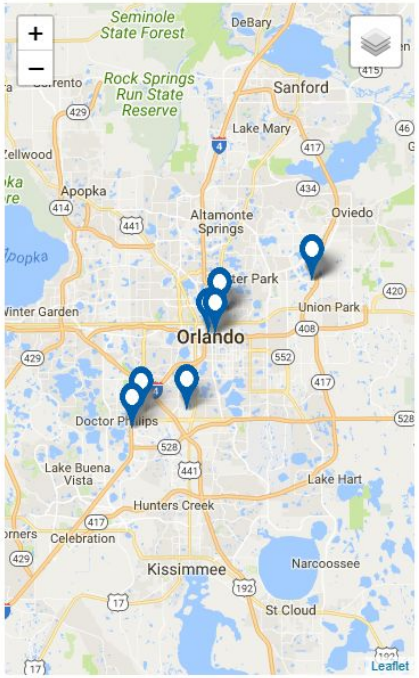
DataKind

E-mail Schedule

DataKind

DataKind | My Activities | Start Point | Optimizer

Map



Schedule Optimizer

Schedule | Daily Hours | Assignments

Optimize

myaddr@example.com E-mail

Monday Time: 4.9 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
CBCCF Team Building	5803 Precision Dr, Orlando, FL, 32819	50 mins	20 mins	⌄
Case A Monthly Visit	2000 Premier Row, Orlando, FL, 32809	60 mins	8 mins	⌃ ⌄
Personal Birthday Party	8625 International Dr, Orlando, FL, 32819	120 mins	10 mins	⌃ ⌄
Finish	4001 Pelee St, Orlando, FL, 32817	-	23 mins	-

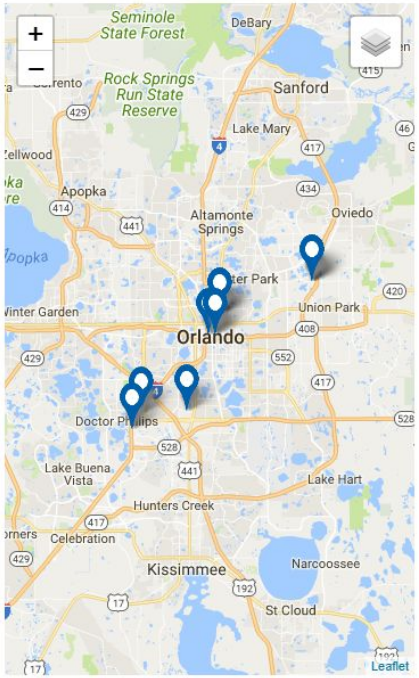
Tuesday Time: 5.6 hrs

Activity	Address	Duration	Travel	Move
Start	4001 Pelee St, Orlando, FL, 32817	-	-	-
Case A Court Visit	22 E Pine St, Orlando, FL, 32801	120 mins	13 mins	⌃ ⌄

Daily Hours

DataKind | My Activities | Start Point | Optimizer

Map



Schedule Optimizer

Schedule | Daily Hours | Assignments

Hours Available:

Monday (8)

Tuesday (8)

Wednesday (5)

Thursday (0)

Friday (5)

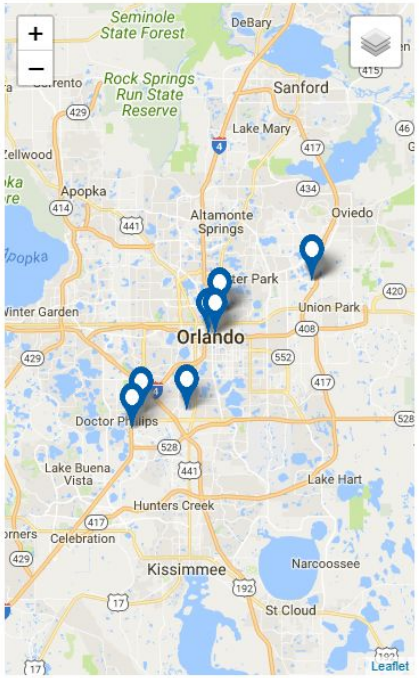
Update | Reset

DataKind

Assignments

DataKind | My Activities | Start Point | Optimizer

Map



Schedule Optimizer

Schedule | Daily Hours | Assignments

Assignments:

Activity | **Assignment**

Add | Clear

Active Assignments:

Activity	Assignment
Case A Monthly Visit	Friday

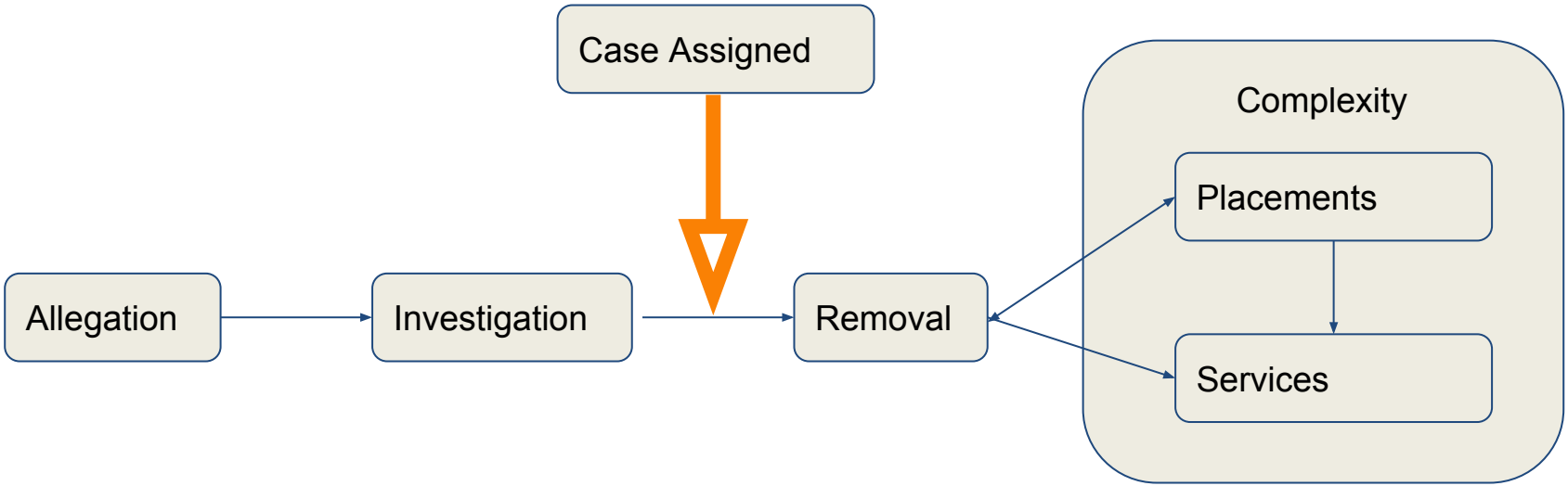
DataKind

Case Complexity

Problem Statement

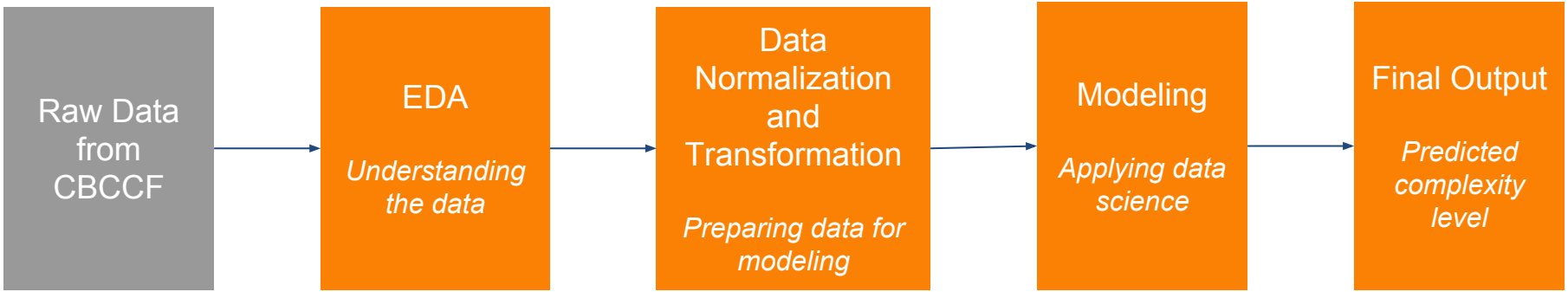
- Current assignment leads to imbalance
- Measuring case complexity will allow managers to balance case assignments
- Give specialized cases to specific people
- Group siblings together or geography

Timeline



DataKind

Our Process

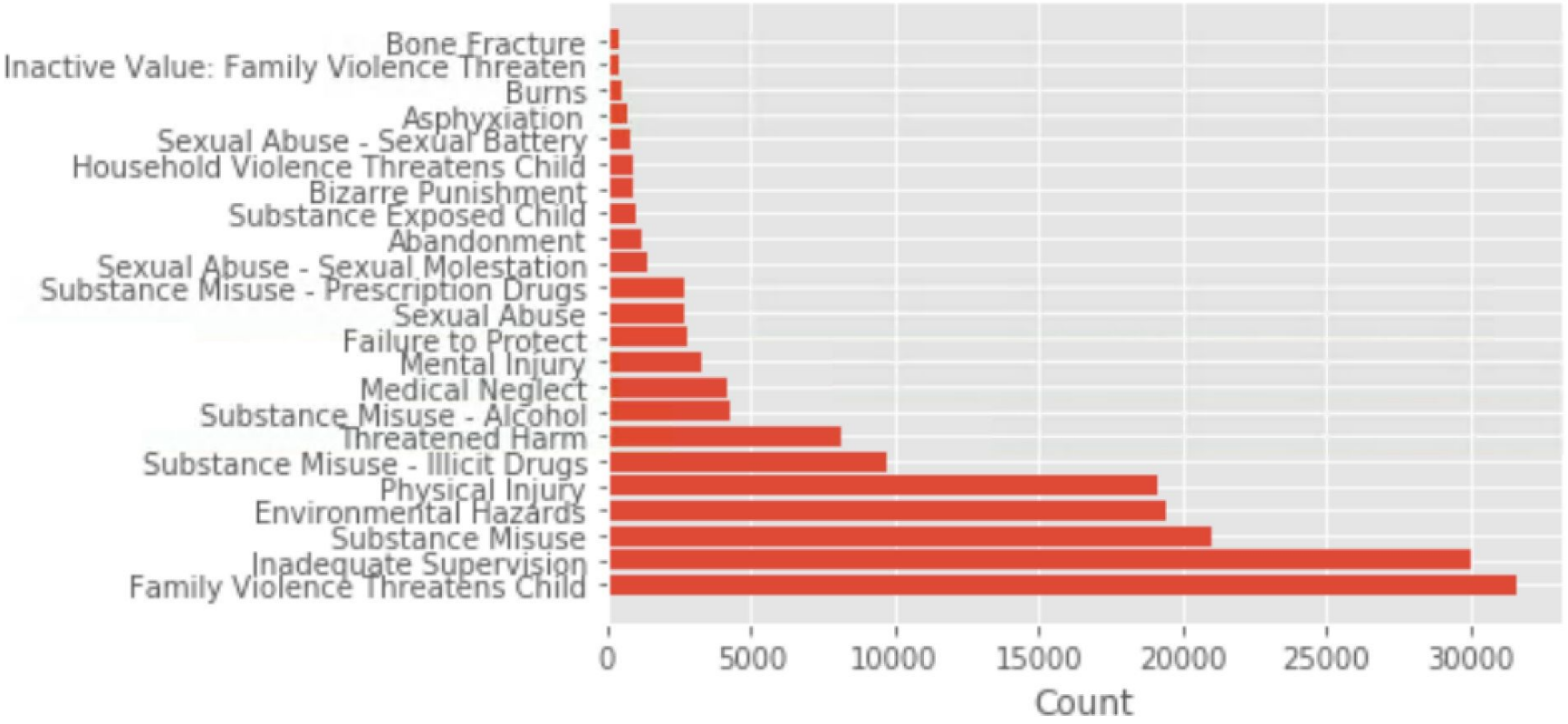


DataKind

Exploratory Data Analysis

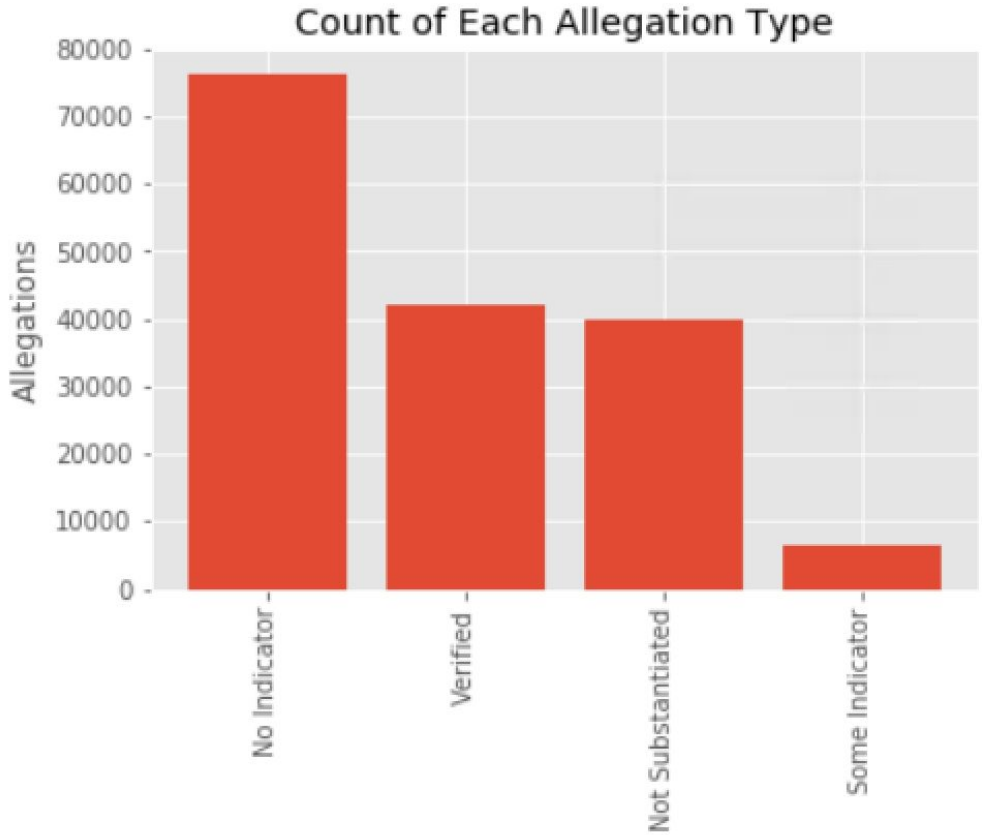
- Volume of Data
- Tables included:
  - Caseworker data
  - Case notes
  - Case notes - person level
  - Allegations
  - Investigations
  - Placements
  - Reasons for Removal
  - Family connections

Allegations



DataKind

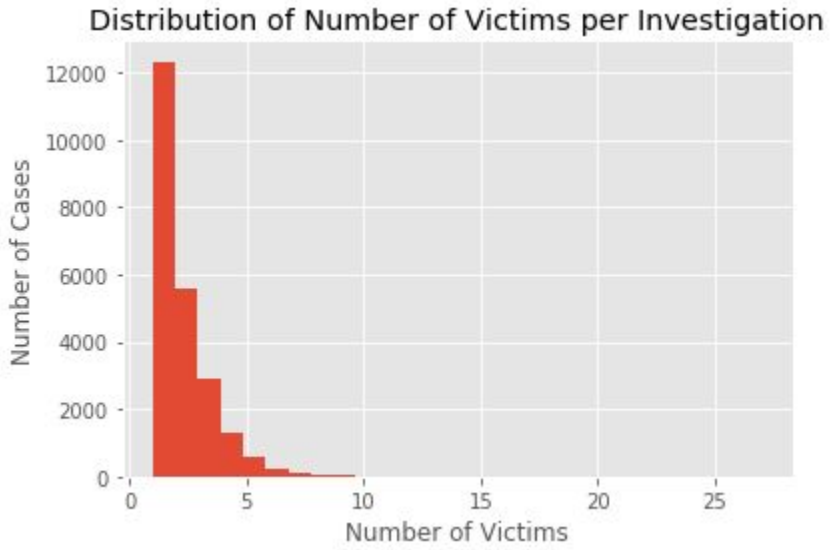
A quarter of allegations end up verified

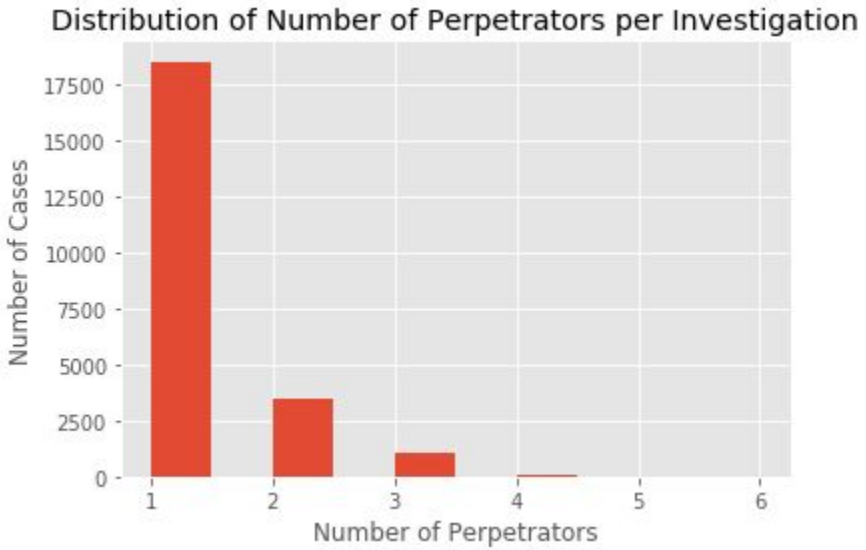


DataKind

Most cases have 1 perpetrator and fewer than 5 victims

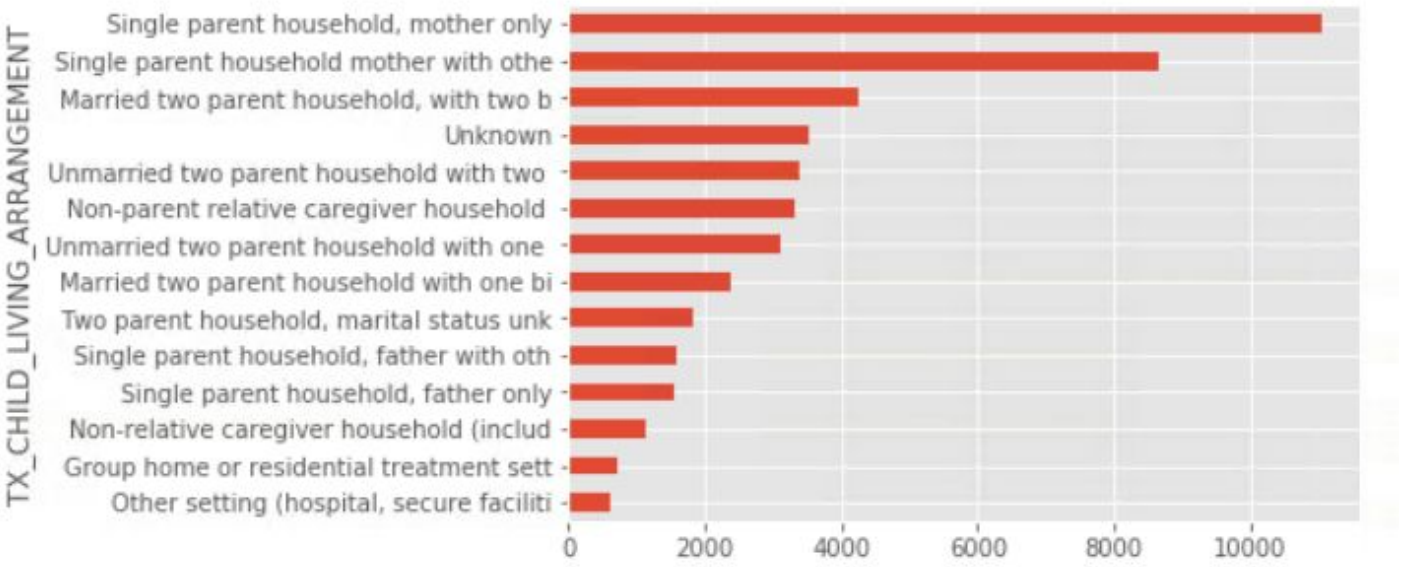
For Seminole County:





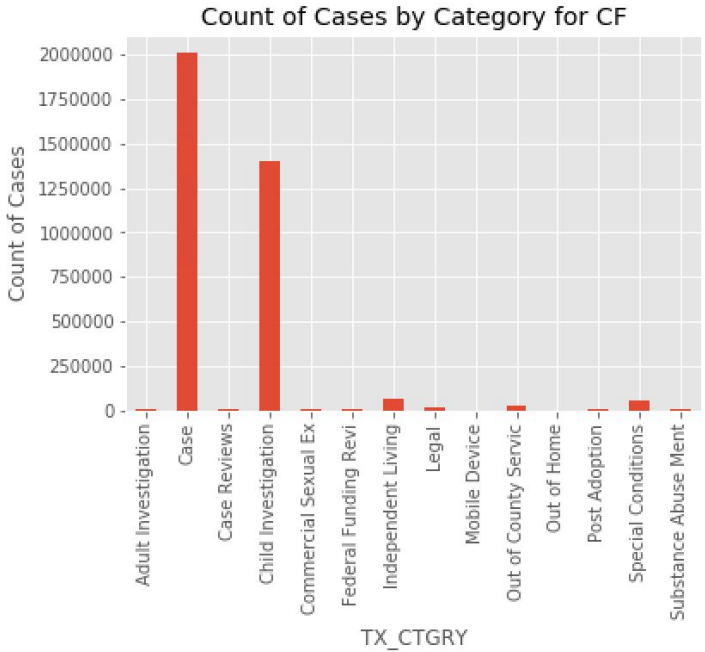
DataKind

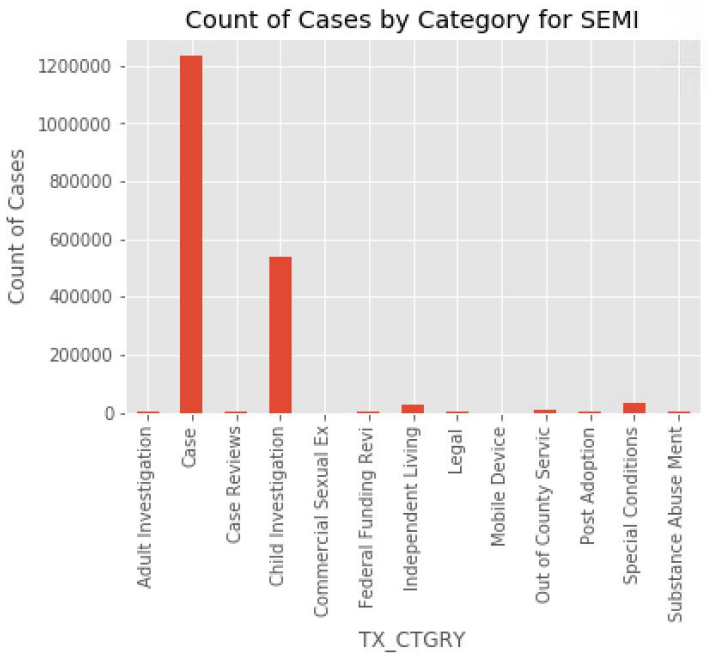
48% of investigations are made in to single parent homes



DataKind

Visualizing Categories

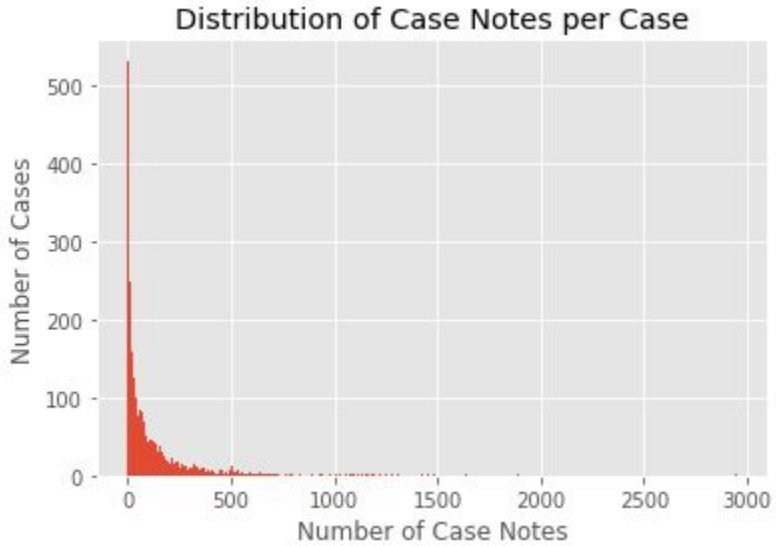




DataKind

The median number of notes/case is 145

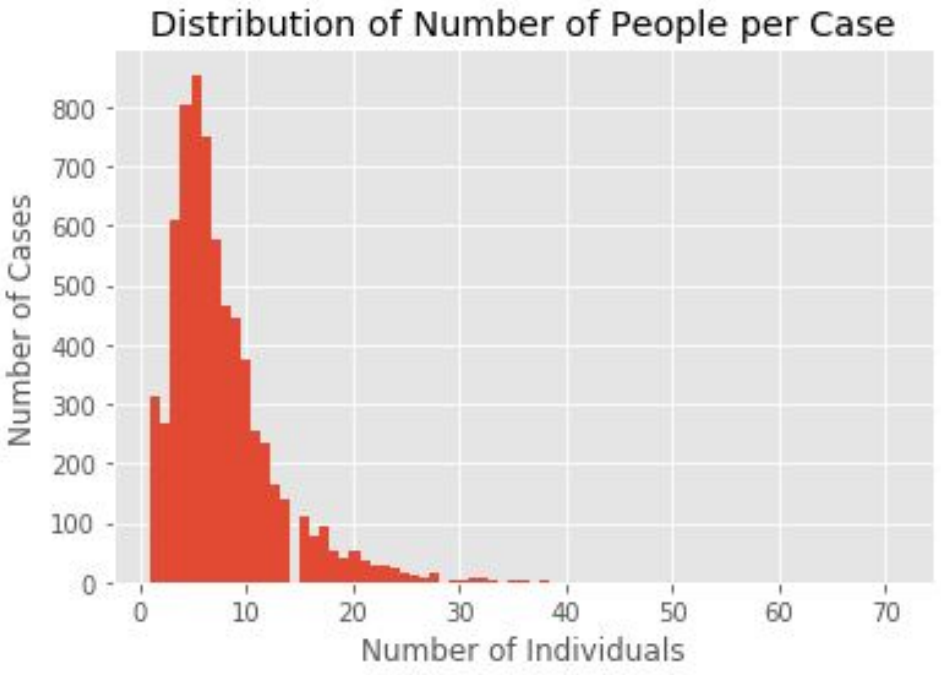
For Seminole County:



DataKind

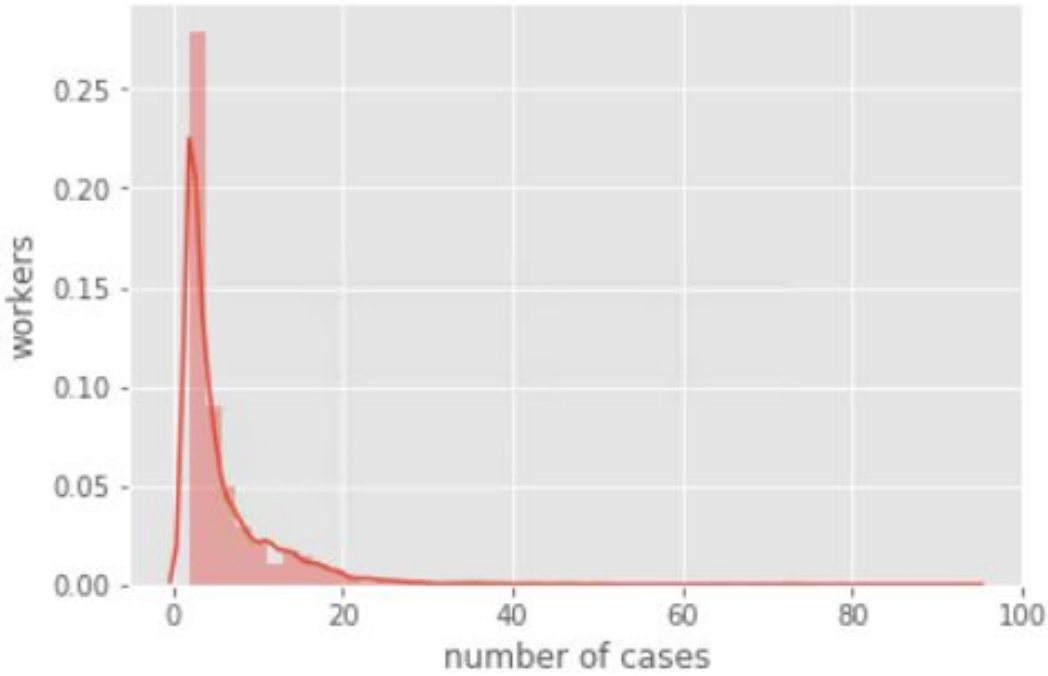
The median number of people/case is 10

For Seminole County:



DataKind

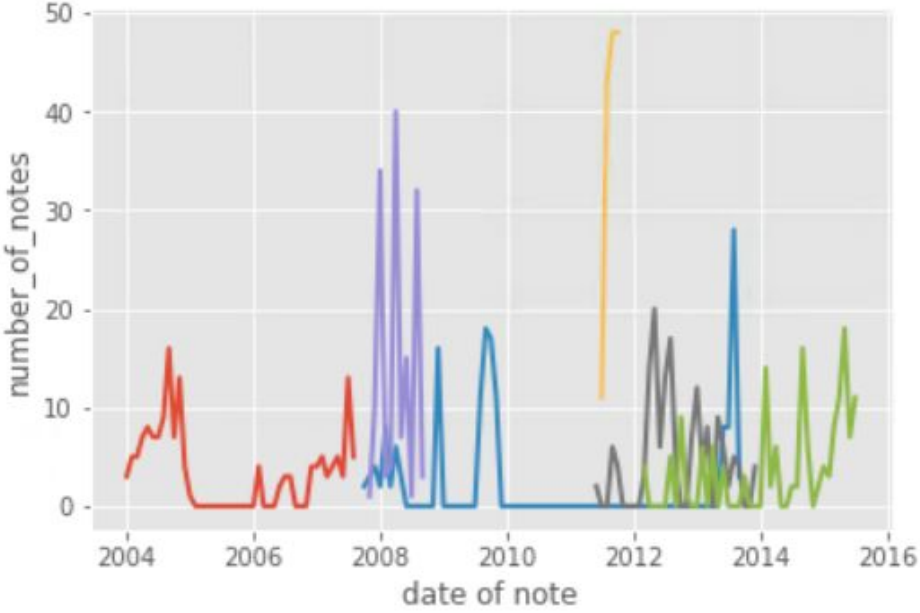
In a month workers log activity for 3 cases



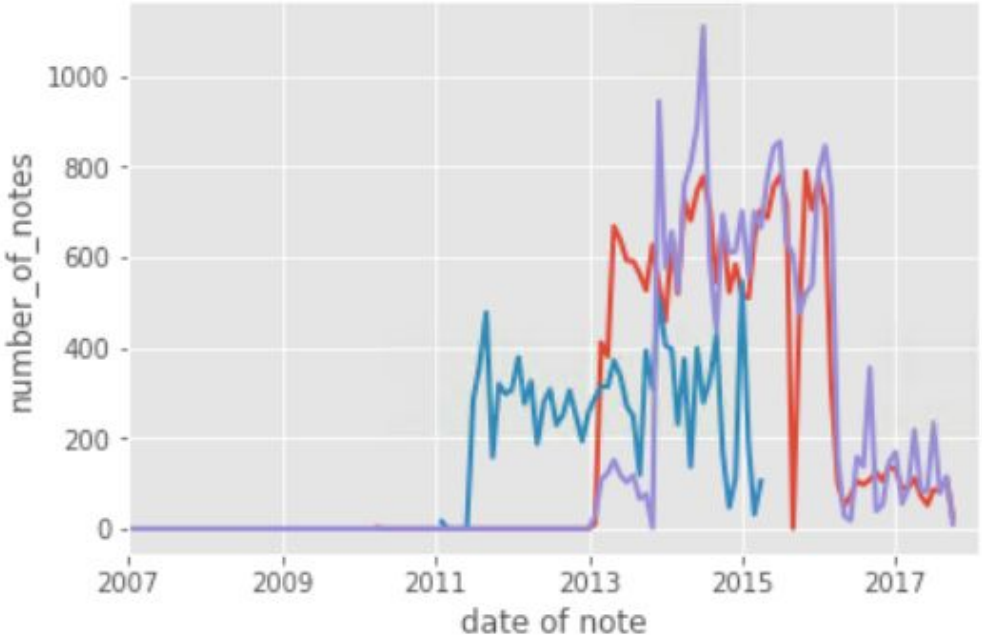
DataKind

Workers log notes in bursts of activity (CF)

Workers with average number of notes (150 notes)



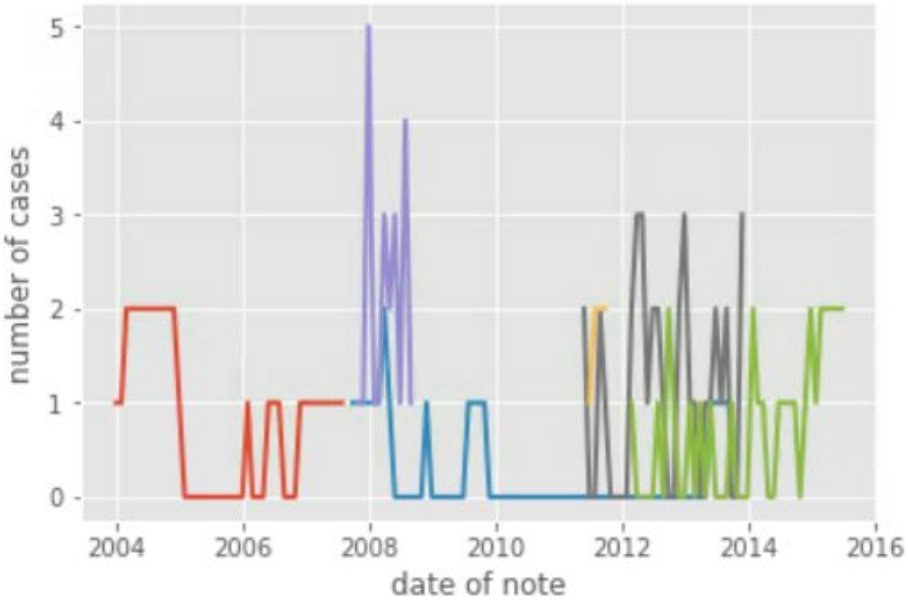
Workers with more than 10000 notes



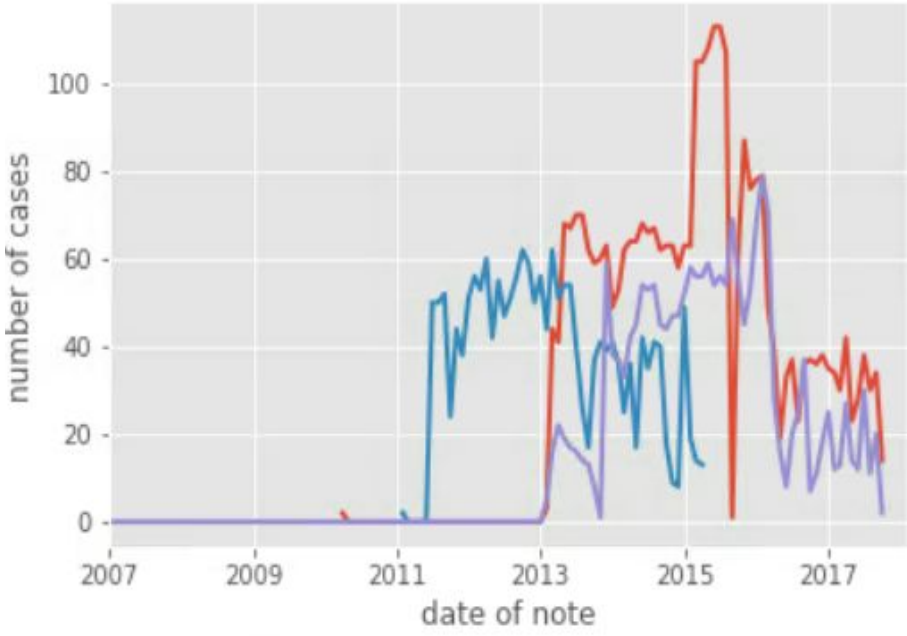
DataKind

There is a huge range in case load (CF)

Workers with average number of notes (150 notes)

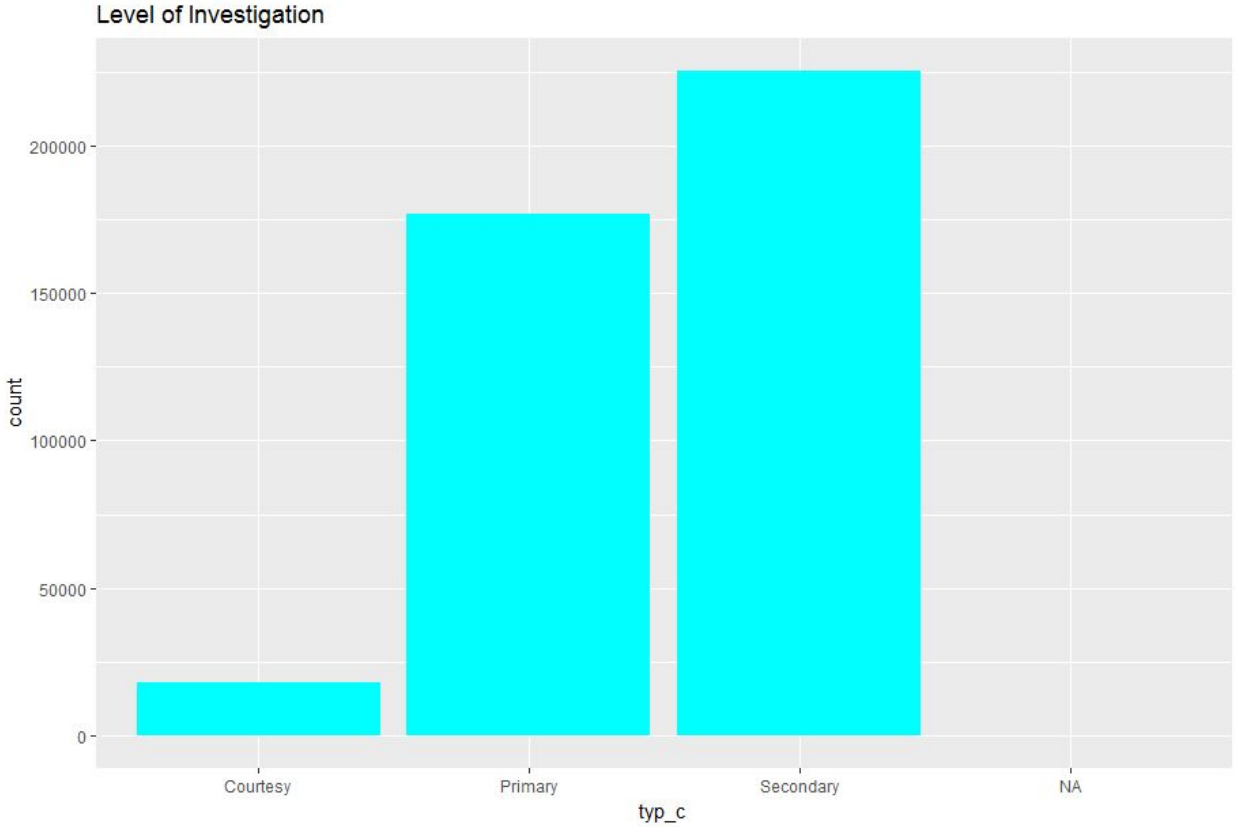


Workers with more than 10000 notes

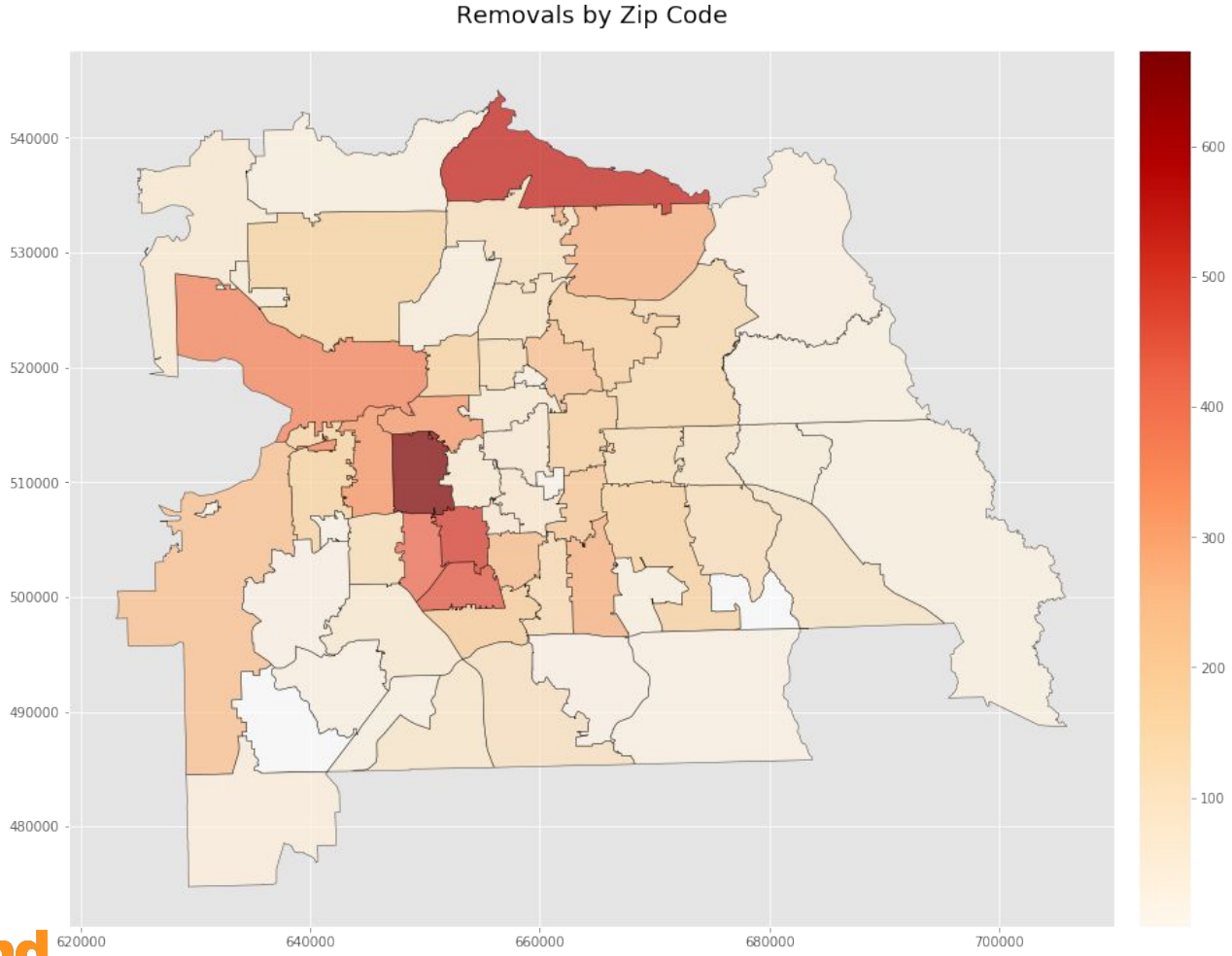


DataKind

Level of Investigation

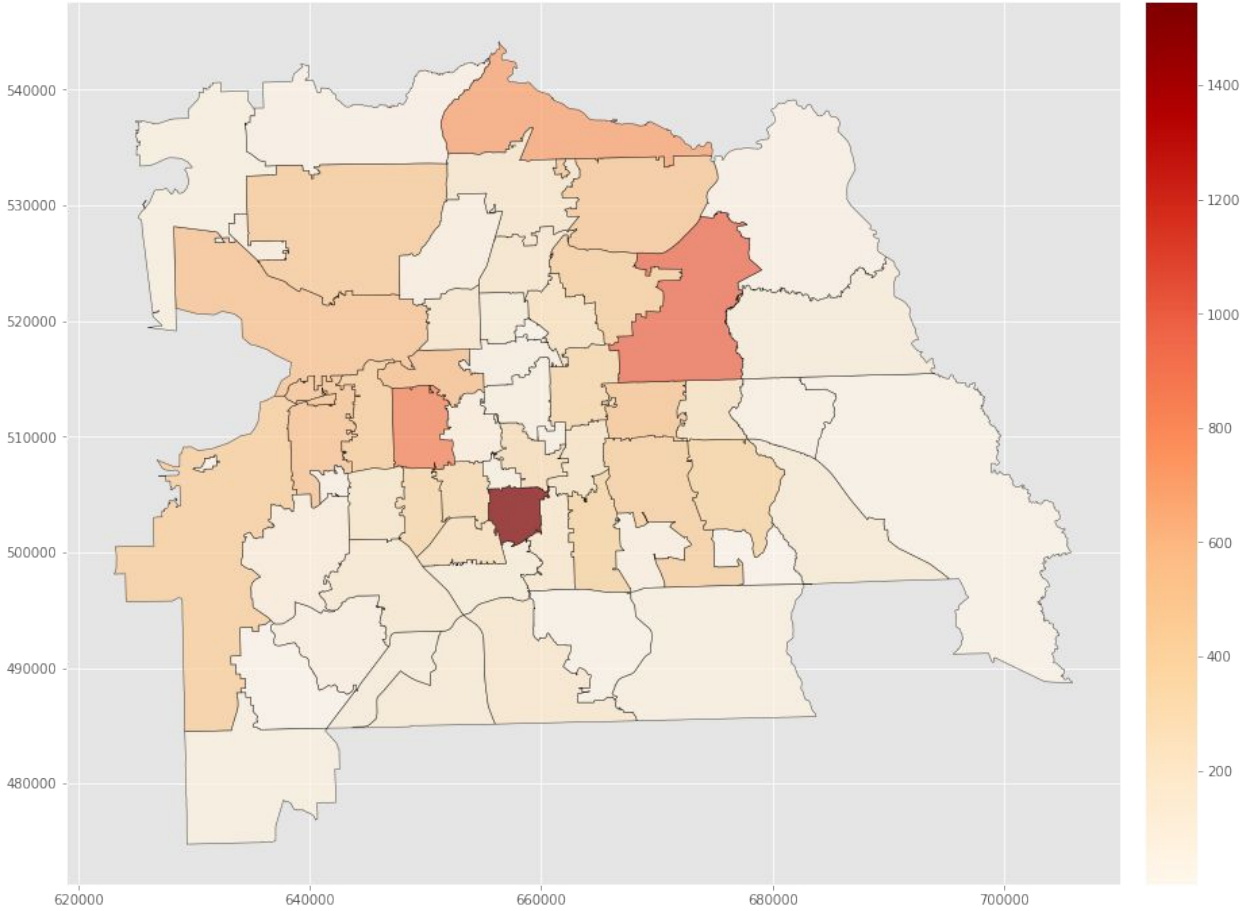


DataKind


Removals by Zip Code
540000
530000
520000
510000
500000
490000
480000
620000
640000
660000
680000
700000
600
500
400
300
200
100

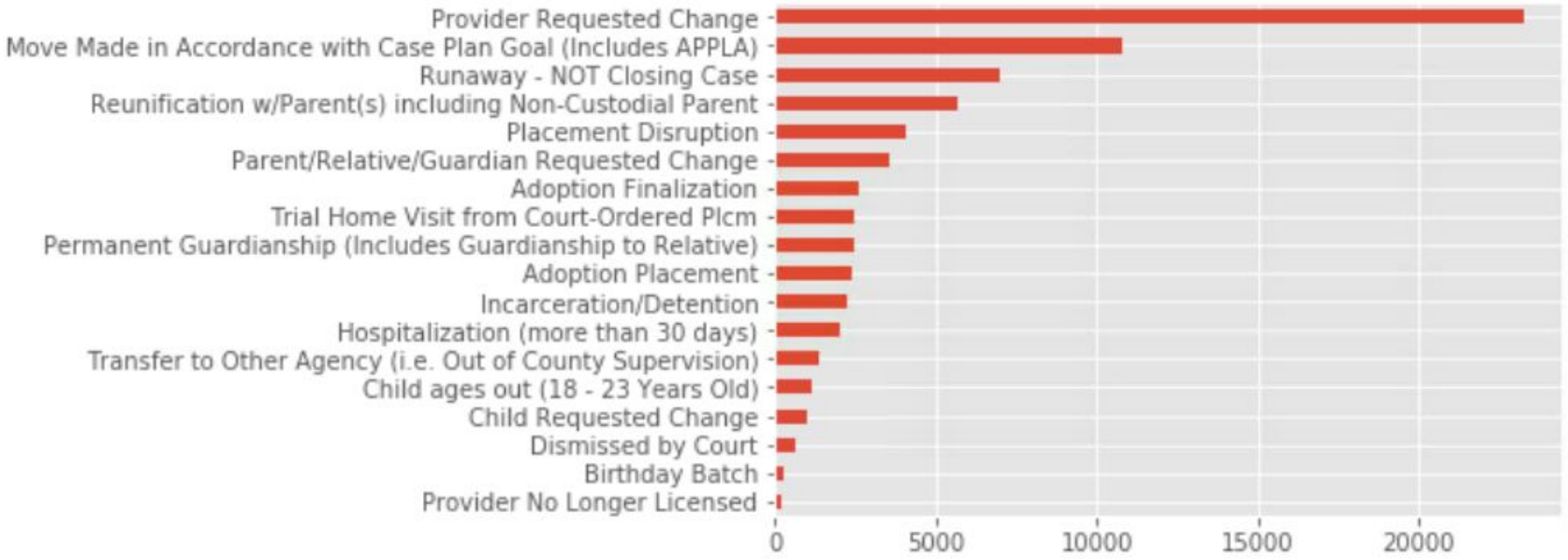
DataKind

Placements by Zip Code



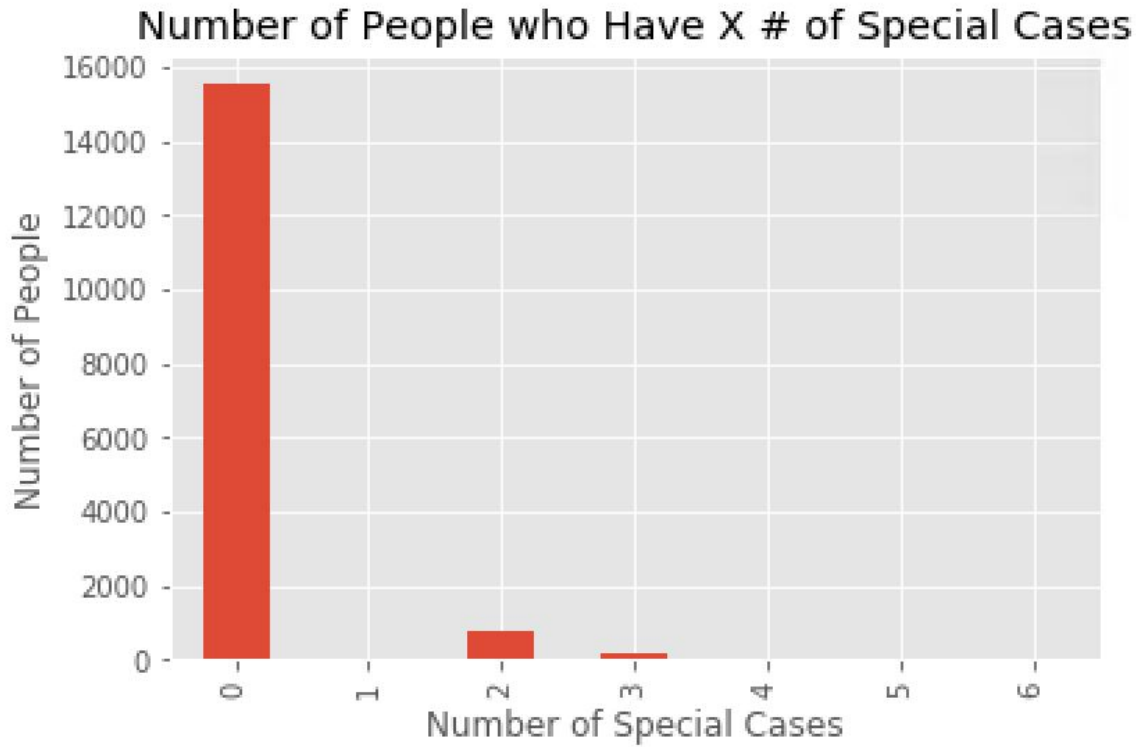
DataKind

Common reasons for why placements end



DataKind

Visualizing Special Cases



DataKind

Questions at the time (2 months in)

- **How can we find when a case was opened (and closed)?**
- **What does the end of a placement mean?**
- **How often do new cases come up?**
- **What time window should we predict complexity in order for it to be relevant during case assignments?**
- With regards to case notes: 1) In a month workers log activity for ~3 cases. Does this seem low? 2) Why are we observing spikes in log activity rather a uniform behavior? Are notes loaded in bulk?

Current Uses of Machine Learning

- Danger Level at Allegation/Investigation
- Avoid High Risk Predictions

Tracing a Case

- Needed to visualize a case to get clear understanding

Fast Facts

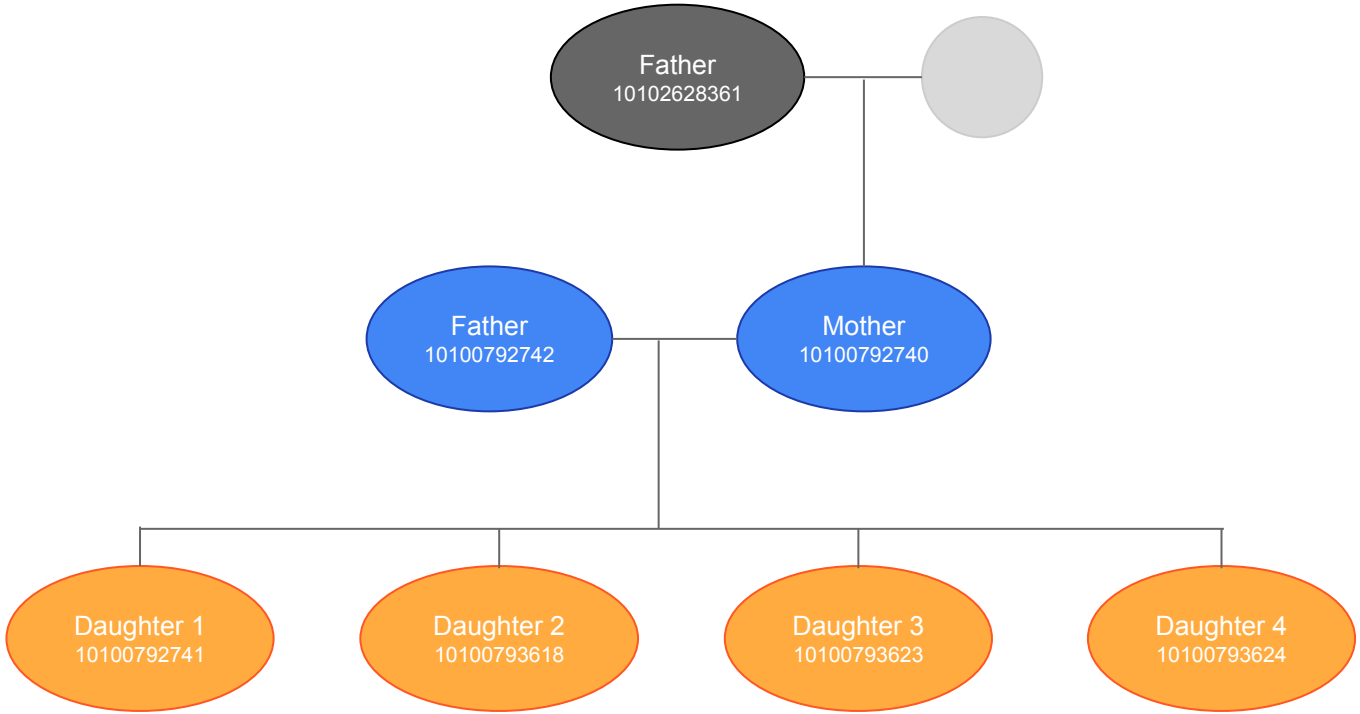
Jan 6, 2009 - June 17, 2014

8 Individuals:
- 4 daughters
- 4 adults (?)
    - 2 parents, 1 maternal grandfather
    - 1 unidentified individual

Allegations:
- Substance misuse
- Threatened harm
- Family violence threatens child

Reasons for Service:
- Inadequate supervision
- Incarcerated parent
- Alcohol abuse by the parent
- Drug abuse by the parent

Family Tree

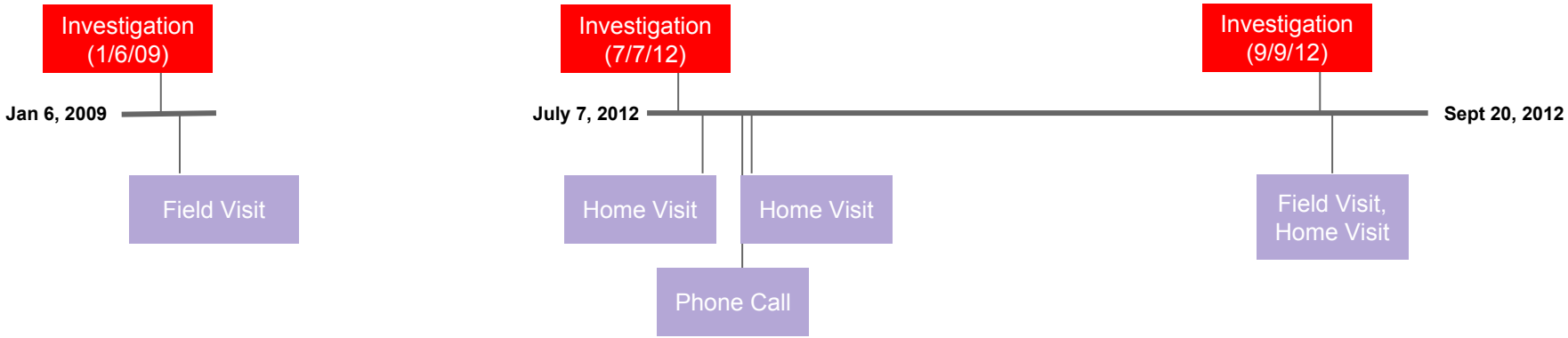

Father
10102628361
Father
10100792742
Mother
10100792740
Daughter 1
10100792741
Daughter 2
10100793618
Daughter 3
10100793623
Daughter 4
10100793624

Fast Facts

14 Caseworkers worked on the case from 7/20/2012 onwards

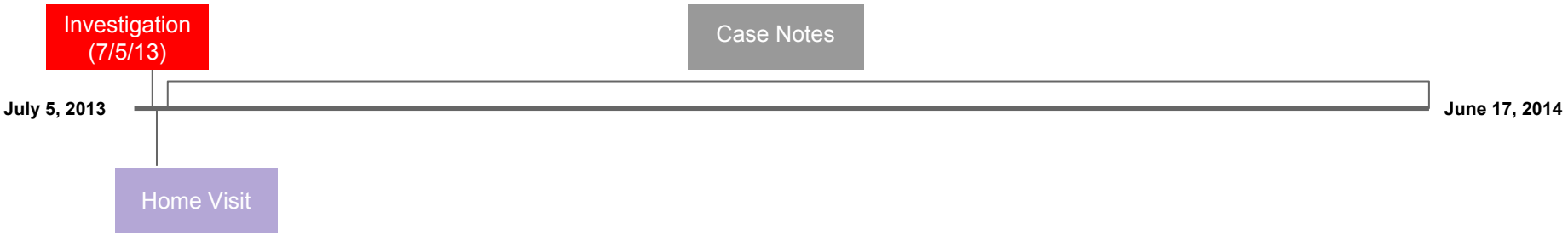
- 4 primary
  - 9922807837: 7/20/12, 11/7/12, 11/9/12
  - 9922808566: 7/20/12 - 11/9/12
  - 9922818164: 7/17/13 - 7/23/13
  - 9922825902: 7/23/13 - 6/30/14
- 10 secondary
  - 9922807837: 7/20/12
  - 9922824272: 7/16/13 - 9/3/13
  - 9922824273: 7/16/13 - 9/3/13
  - 1180040000: 7/16/13 - 6/30/14
  - 9922767926: 7/16/13 - 6/30/14
  - 9922766481: 7/16/13 - 6/30/14
  - 9922790794: 7/19/13 - 6/30/14
  - 9922797904: 10/3/13 - 11/5/13
  - 9922823537: 11/18/13 - 11/27/13
  - 9922711699: 11/26/13 - 6/30/14
- 1 Out of County
  - 9922789900: 11/8/13 - 11/11/13

Timeline of the Case

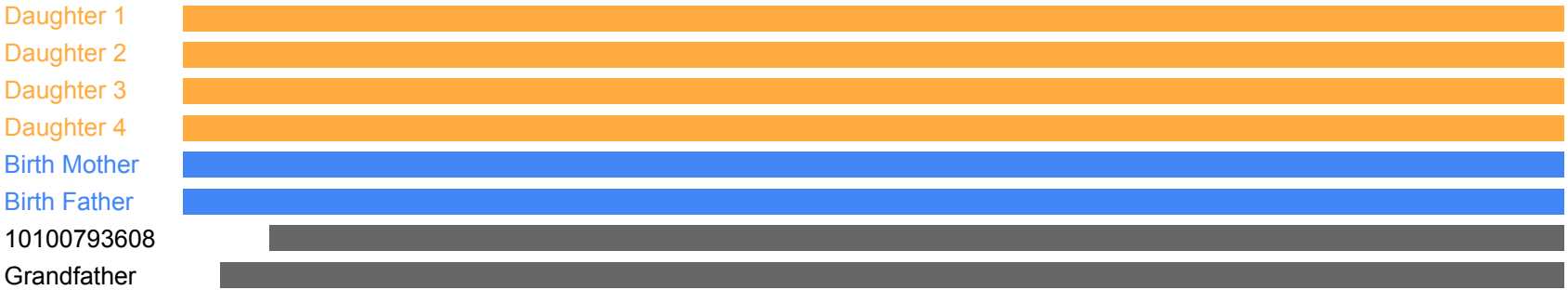


Persons Involved:

Timeline of the Case (cont'd)



Persons Involved:



1/6/09 – Investigation # 1

Living Arrangement: married two parent household, with two b
CPS Type: In-Home
Response Priority: 24 hours
Allegations and investigated outcome:

Child	Allegation	Perpetrator	Outcome
Daughter 1	Substance Misuse	0	Some indicator
	Threatened Harm	0	No indicator
Daughter 2	Substance Misuse	0	Some Indicator
Daughter 3	Substance Misuse	0	Some Indicator
Daughter 4	Substance Misuse	0	Some Indicator

1/6/2009 - 1/10/2009

1 Investigation

1 Field Visit

7/7/12 – Investigation # 2

Living Arrangement: married two parent household, with two b
CPS Type: In-Home
Response Priority: 24 hours
Allegations and investigated outcome:

Child	Allegation	Perpetrator	Outcome
Daughter 2	Family Violence Threatens Child, Substance Misuse	0	Not Substantiated
Daughter 3	Family Violence Threatens Child, Substance Misuse	0	Not Substantiated
Daughter 4	Family Violence Threatens Child, Substance Misuse	0	Not Substantiated

7/7/2012 - 9/13/2012

1 Investigation
1 Home Visit to the Child's Current Residence
1 Home Visit to the Parent/Caregiver
1 Telephone Contact
1 Note to File

9/19/12 – Investigation # 3

Living Arrangement: Single parent household, father only
CPS Type: In-Home
Response Priority: 24 hours
Allegations and investigated outcome:

Child	Allegation	Perpetrator	Outcome
Daughter 2	Substance Misuse	Mother	Verified
Daughter 3	Substance Misuse	Mother	Verified
Daughter 4	Substance Misuse	Mother	Verified

9/19/2012 - 9/20/2012

1 Investigation

2 Field Visits

1 Home Visit to the Child’s Current Residence

7/5/13 – Investigation # 4

Living Arrangement: Single parent household, father only
CPS Type: In-Home
Response Priority: 24 hours
Allegations and investigated outcome:

Child	Allegation	Perpetrator	Outcome
Daughter 2	Substance Misuse	Mother	Verified
Daughter 3	Substance Misuse	Mother	Verified
Daughter 4	Substance Misuse	Mother	Verified

7/5/2013 - 6/17/2014

1 Investigation

1 Home Visit to the Parent/Caregiver

46 Case Notes missing tx_type

Minimum Viable Product

- Simplest iteration possible
- See feasibility of machine learning (is this possible?)

Recap

- How do we measure complexity?
- Is it possible to give specialized cases to certain people?
- Is it possible to take geospatial information into account?

How to measure complexity

- Initial Idea: Predicting End Date

Problem Setup

How do we define “complexity”?

Our approach is to use activity as a proxy for a case’s complexity.

- The activity rate is defined as the number of case notes entered in a 30 day period.
- We then convert the activity rate into a duration in hours using the estimated duration data provided by CBCCF

Why use an activity rate?

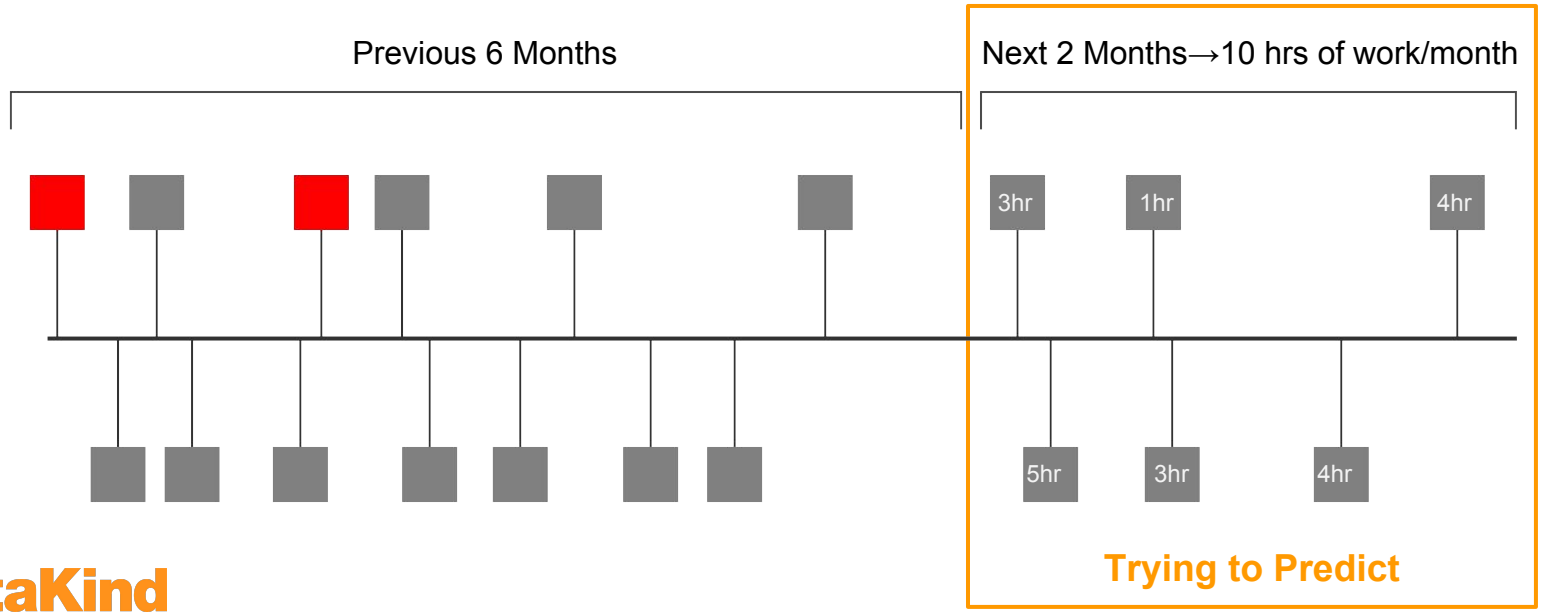
The advantages of using a activity duration:

- Easy to understand
- Actionable
- Less influenced by a caseworker’s ability
- Somewhat independent of the number of caseworkers assigned to a case, since we can’t fully segment out work done by individual case workers on a case

Problem Setup

How are we calculating complexity? (What we are predicting...)

For each case, we will be using data attributes from the previous 6 months of the case to predict the average level of activity for the next 2 months.



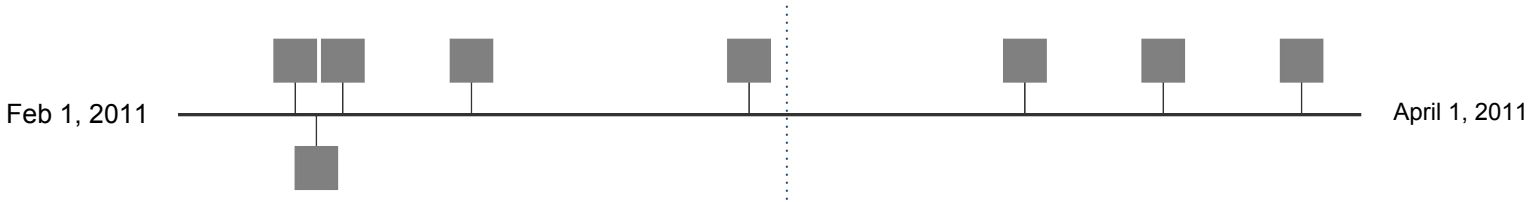
DataKind

Model predicts: average hours/month

Counting Average Activity

- Calculated the average number of hours spent in the seventh and eighth months of each case
  - This is what we will be trying to predict

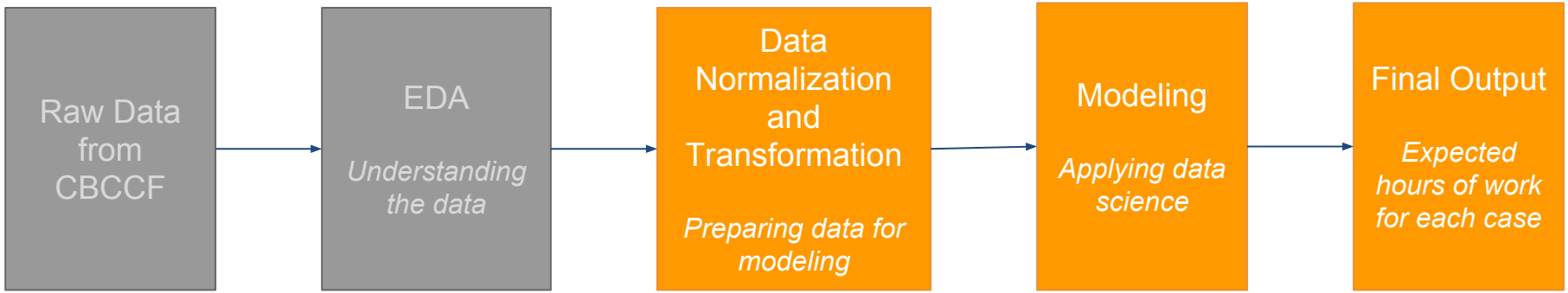
An example: **Case 10001186814**



Feb 1, 2011 - March 2, 2011	March 3, 2011 - April 1, 2011	Average hours
5 notes, 8.5 hours spent	3 notes, 3 hours spent	5.75

DataKind

Our Process



DataKind

Representative case dataset

Determining Eligible Cases:

- At least 3 months long but no longer than 18 years
- Minimum of 5 case notes
- Caseworkers with 10 or more handled cases
- Total of 8,907 cases from CBC jurisdiction included

Activity in the representative case dataset

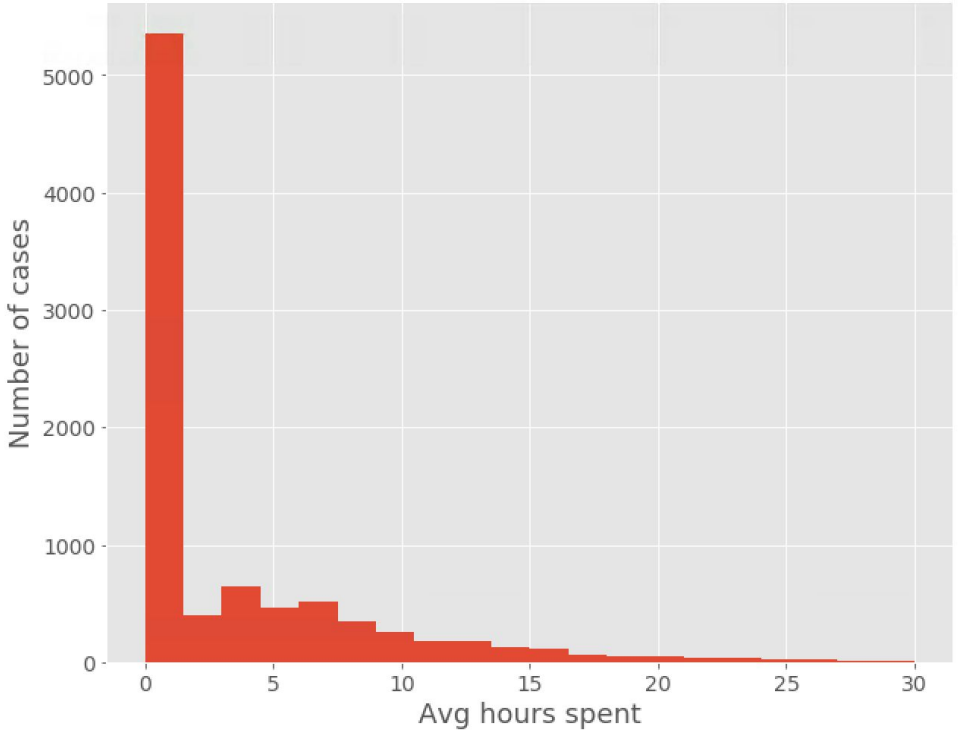
Mean: 3.56 hours

Median: 0.0 hours

(5258 cases)

Max: 159. Hours

(Case id: 10001763991)



DataKind

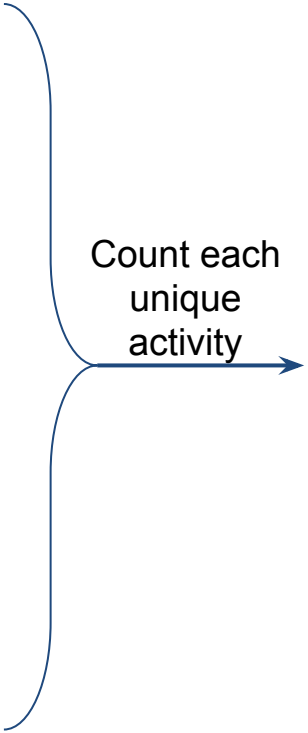
Data we normalized and aggregated

Investigations: • Living arrangement • Access type • CPS type • Determination • Response priority	**Placements**: • Filter to placement -related services • Number of placements • Number of placement zip codes • Placement duration	**Removals**: • Manner of removal • Family structure • Discharge reason • Reason for service
Reason for Service: • Reasons for why child is considered at-risk	**Notes:** • Note category • Type of work being done	

DataKind

Count vectorizing: what the models sees

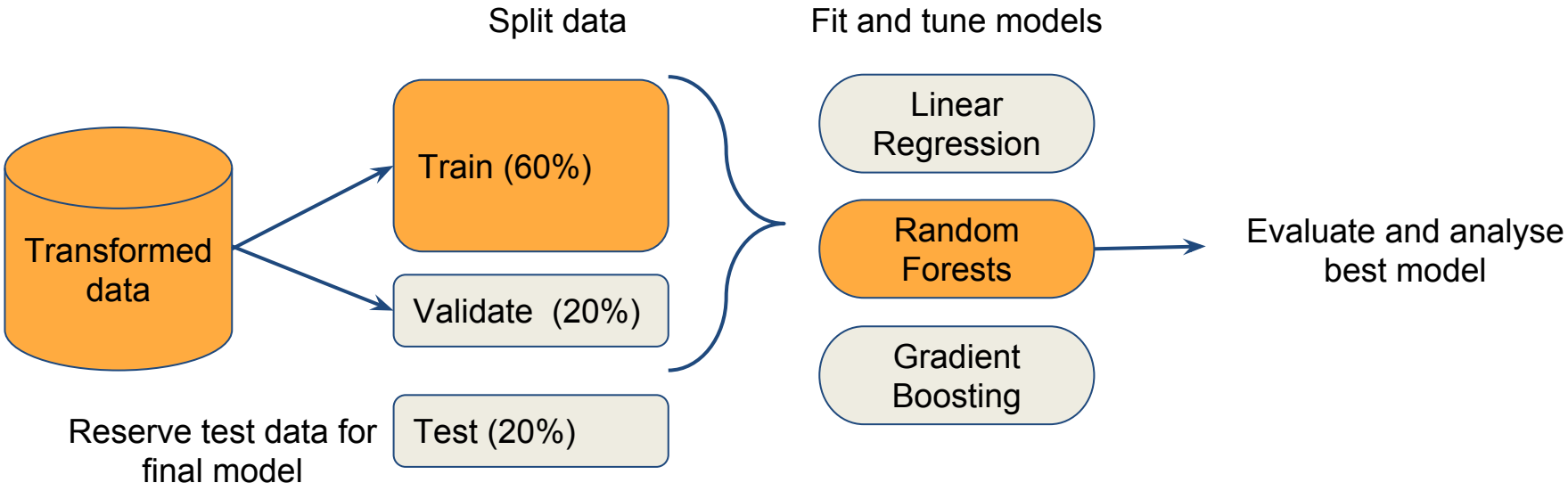
Activity Type	Date
Investigation	7/7/12
Home Visit	7/8/12
Telephone Contact	7/16/12
Home Visit	7/17/12
Home visit	9/13/12
Investigation	9/19/12
Placement	9/20/12
Placement	9/25/12



Investigation	Home Visit	Telephone Contact	Placement
2	**3**	**1**	**2**

DataKind

Modeling Process



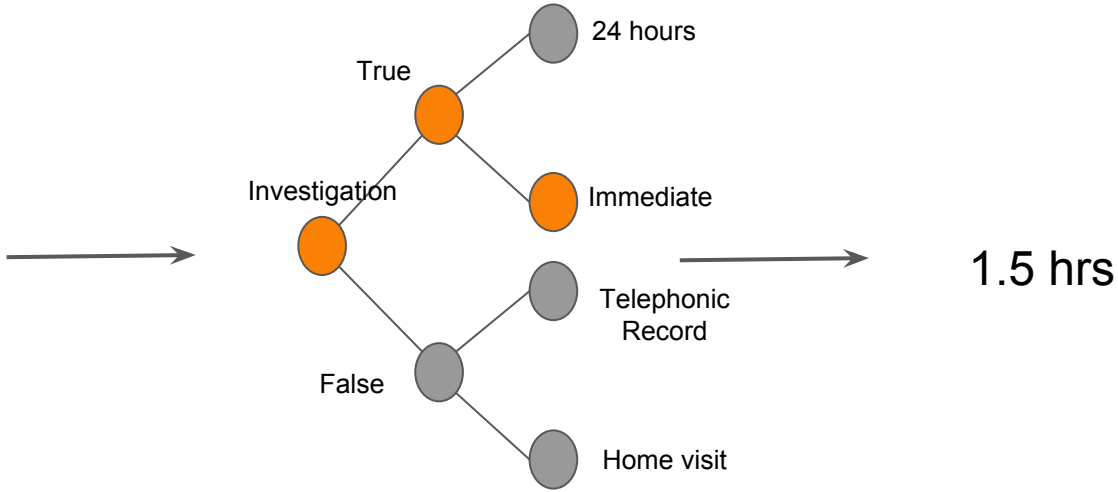
Mean absolute error (MAE) : actual hours spent - predicted hours

DataKind

Decision tree

Decision Tree Classifiers: Decision Tree Classifiers are used to predict outcome of an observation. The model maps information about a specific observation to a specific outcome. Such models are popular because of their ease of interpretation and simplicity of application. The model can be thought of as a series of if/then statements that allow for prediction of an observation belonging to a specific group/outcome. The if then statements attempt increase homogeneity (creating like groups) within groups.

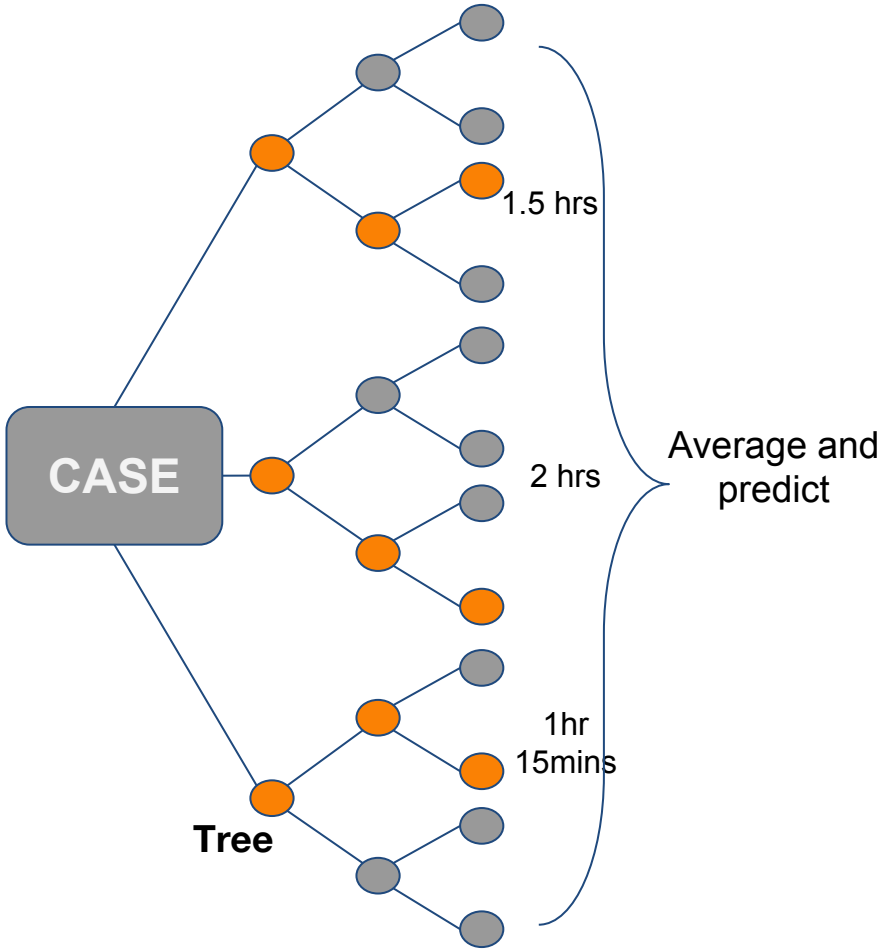
Investigation	Home Visit	Telephone Contact	Placement
2	3	1	2



DataKind

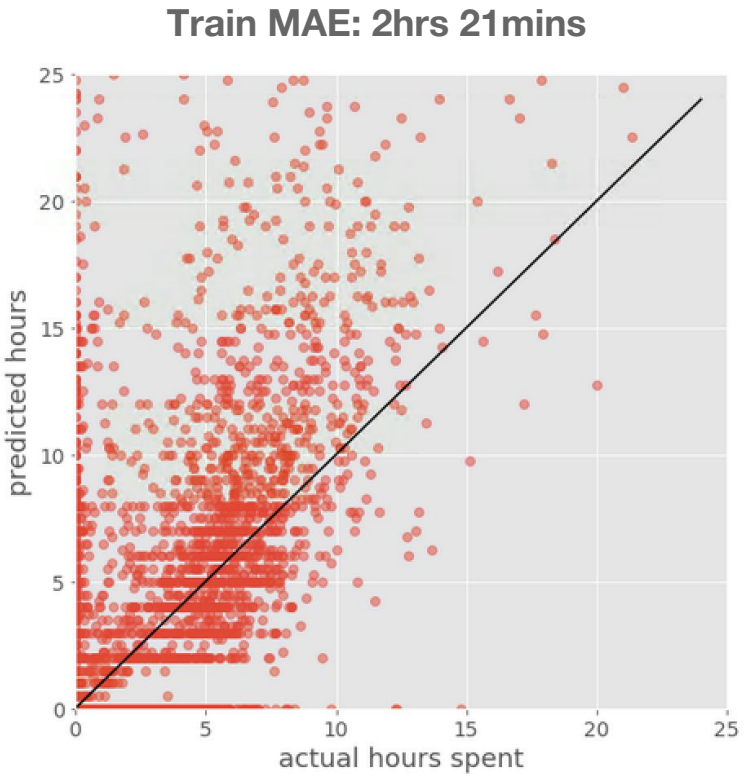
Random forests

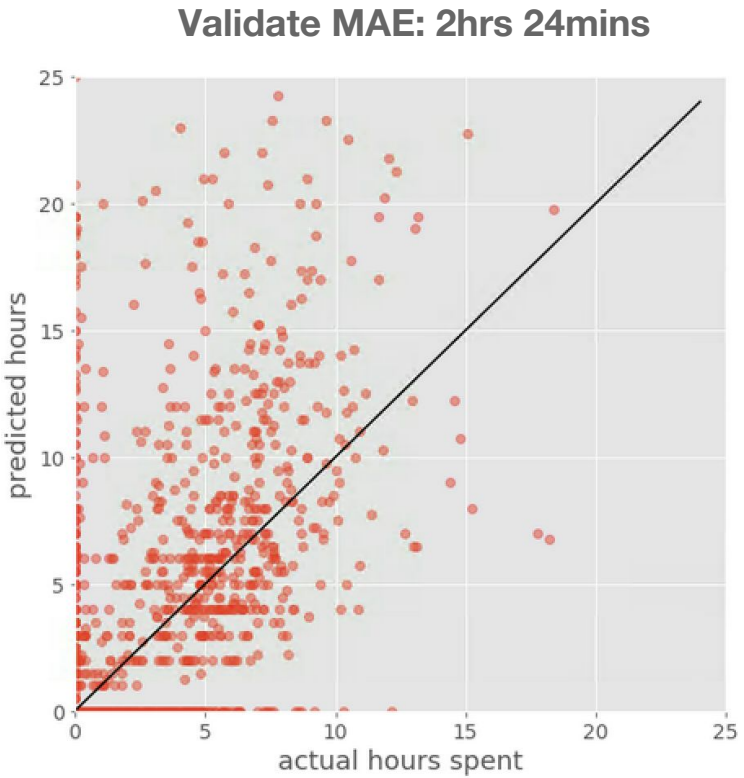
- Random Forest Classification uses the Classification Tree approach but instead of estimating a single tree, multiple trees are created (i.e. tree -> forest).
- This algorithm is referred to as an “ensemble” method as it brings together multiple models to predict a single outcome. Each observation traverses each tree and each tree “casts a vote” the proportion of votes are used to create prediction.
- This algorithm has strong predictive power, can handle missing or incomplete data and is relatively easy to implement.
- One downside is that the model logic is difficult to interpret, but one can generate a list of the most important features in the model.



DataKind

On average we have an error of 2hrs 24mins

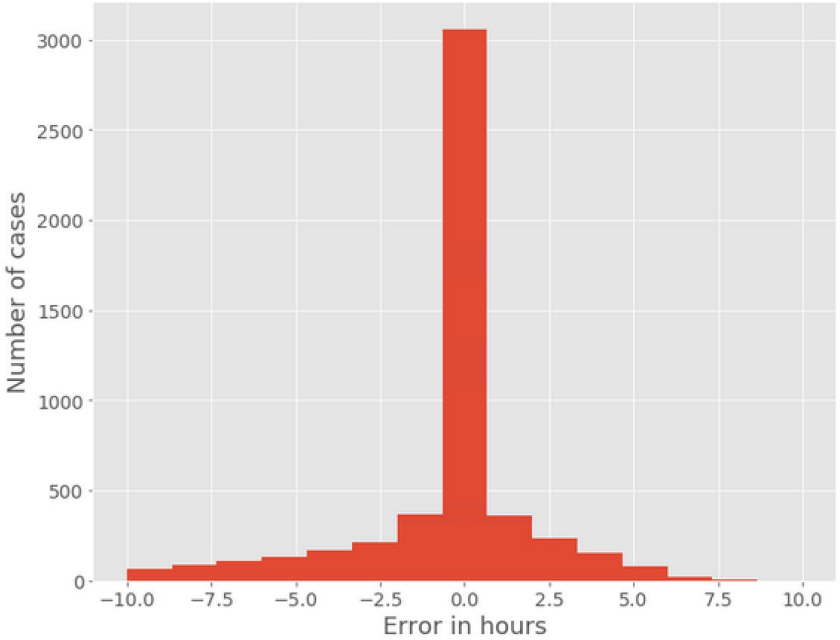




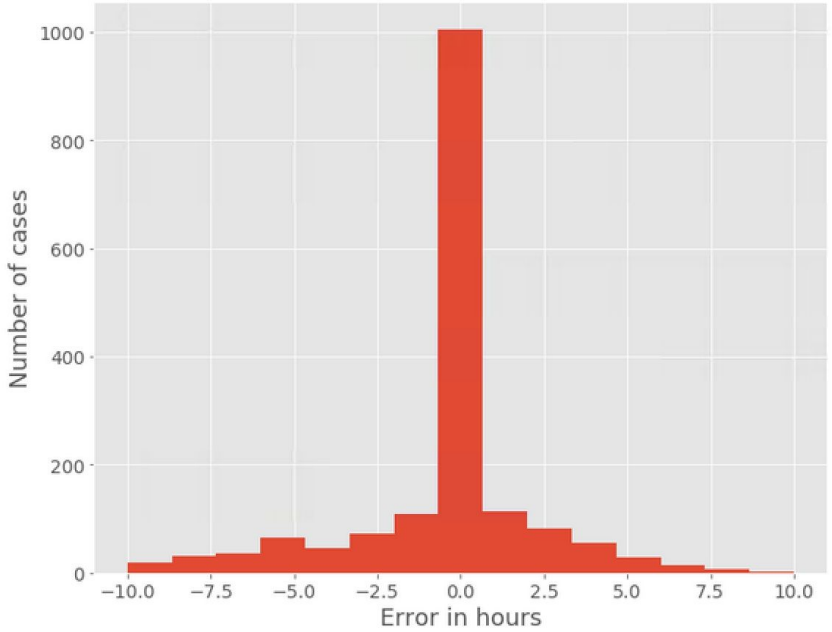
DataKind

Frequency distribution of errors

Train MAE: 2hrs 21mins



Validate MAE: 2hrs 24mins



DataKind

20 most important features

Investigations:

- Child Investigation Initial Face-to-Face

Placements:

- Count of placements
- No of_distinct_zips
- total_days_per_placement
- placement_cat_Foster Home
- placement_cat_Relative

Removals:

Reason for Service:

- Closing-Services
- Child Intake
- In-Home

Notes:

- Case Note to File - Legal
- Case Field Visit - Other
- Case Note to File - General
- Case Telephone Contact
- Case Court - Judicial
- Case Reviews - Supervisory
- Case Home Visit - Child's Current Residence

Notes:

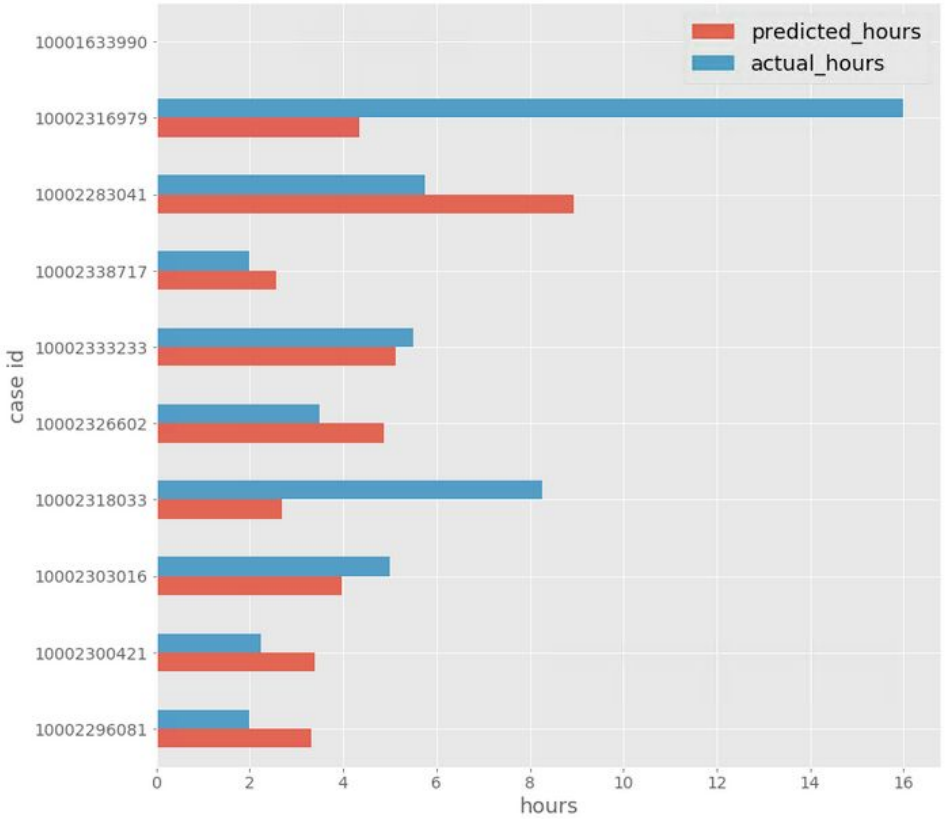
- Case Service Provider Contact
- Case Office Visit
- Case Other
- Case Home Visit - Parent/CG Removed From

DataKind

Example use case

- Caseworker with 9922849324 was assigned to 10 cases as a primary starting between January 2017 and April 2017
- On the 1st, Aug 2017 a Supervisor wanted to estimate how many hours would this case worker likely work in the next two months.
- The model was used to predict the hours they would spend on 1st, Aug 2017 for the next two months.
- At this point in time every case had been active for at least 3 months
- We then analyzed what features the model saw for each case

What did we predict?



DataKind

Case 3233, Predicted: 5.1hrs, Error: 0.4 hrs

Investigations: • Single parent household, mother only • 24 Hours • Child Investigation Commencement (Initial)	**Placements**: • placement_cat_Relative, • placement_flg_Y, • Total_days_per_placement • no_of_distinct_zips	**Removals**:
Reason for Service: • Closing-Services • Child Intake • In-Home • Closing-Open Ongoing Case Management Ser,	**Notes:** • Case Telephone Contact • Case Note to File - Legal • Case Note to File - General • Child Investigation • Telephone Contact • Child Investigation Note • Child Investigation Home Visit - Parent/Caregiver, •	**Notes:** • Case Court - Judicial, • Case Visitation - Parent/CG Removed • Case Service Provider Contact • Case Home Visit - Child's Current Residence

DataKind

Case 8033, Predicted: 8.25hrs, Error: 5.5 hrs

Investigations: • Single parent household, mother only • Immediate,	**Placements**: • placement_cat_Medical - Hospitalization • placement_cat_Relative • placement_flg_Y • Total_days_per_placements • no_of_distinct_zips	**Removals**:
Reason for Service: • Child Intake, In-Home, • Closing-Open Ongoing Case Management Ser	**Notes:** • Child Investigation Home Visit - Child's Current Residence • Child Investigation Field Visit • Case E-Mail Contact • Child Investigation Commencement (Initial) • Case Note to File - General	**Notes:** • Child Investigation Telephone Contact • Child Investigation Note to File - General • Case Supervisory Consult • Case Home Visit - Child's Current Residence

DataKind

Case 6979, Predicted: 8.9hrs, Error: -3.2 hrs

Investigations: • Single parent household mother with othe • 24 Hours, Immediate	**Placements**:	**Removals**:
Reason for Service: • Child Intake, • Closing-No Services, • Closing-Open Ongoing Case Management Ser • fl_drg_abs_prnt	**Notes:** • Child Investigation Field Visit • Case E-Mail Contact • Case Office Visit • Case Visitation - Relative, • Case Telephone Contact • Case Note to File - General, Child Investigation	**Notes:** • Child Investigation Home Visit - Parent/Caregiver • Case Visitation - Parent/CG Removed From • Case Service Provider Contact, Case Home Visit - Child's Current Residence,

DataKind

Post-MVP Model Improvements

Sampling changes

- Include activity from the middle and end of cases, not just the beginning

Model tuning

- Make adjustments to improve model performance

Model interpretability

- Shed light on why the model made the predictions it did

DataKind

Including activity from the middle and end of cases

Before:

All activity was drawn from the first six months of a case

Now:

⅓ of sample is from the first six months

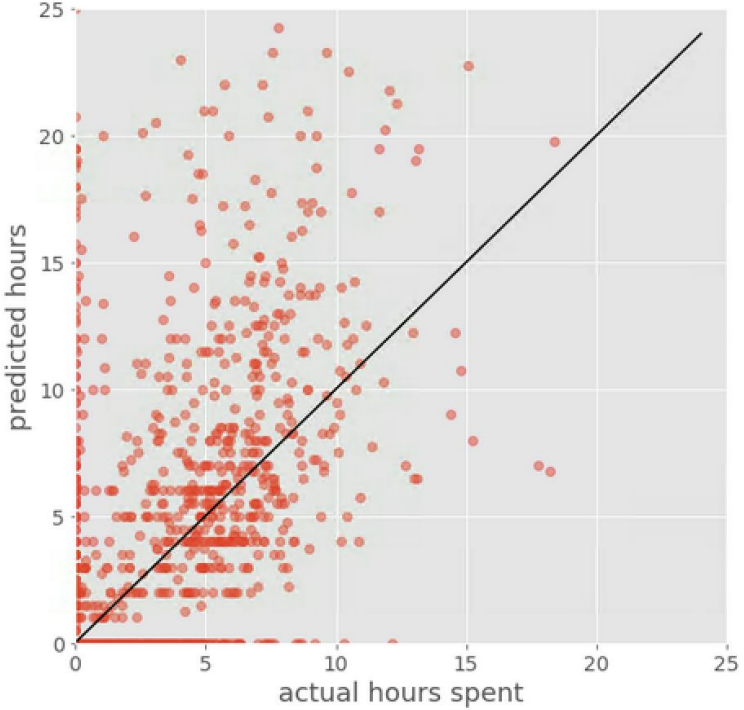
⅓ of sample is from the last six months

⅓ of sample is randomly drawn from the middle

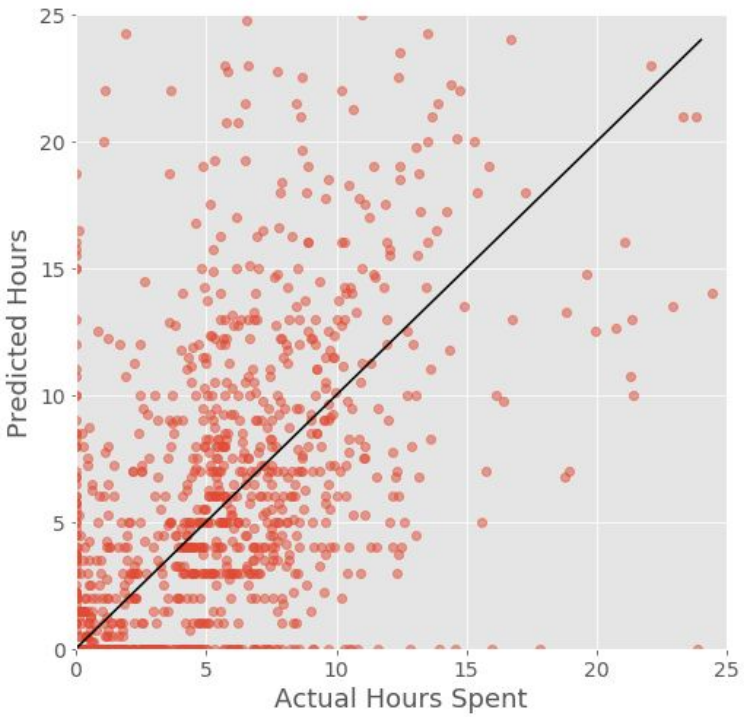
DataKind

Improving Model Accuracy

MVP Model MAE: 2hrs 24mins



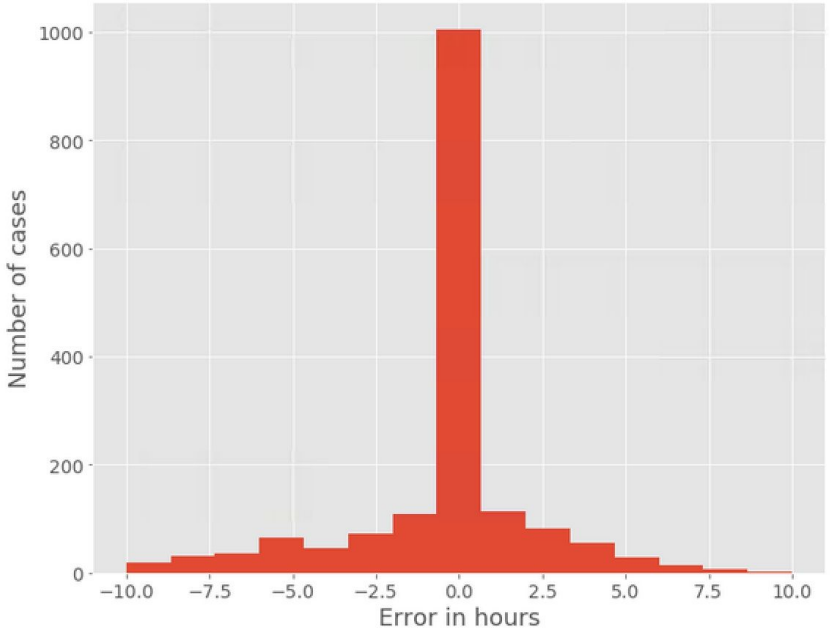
Current Model MAE: 4hrs 30mins



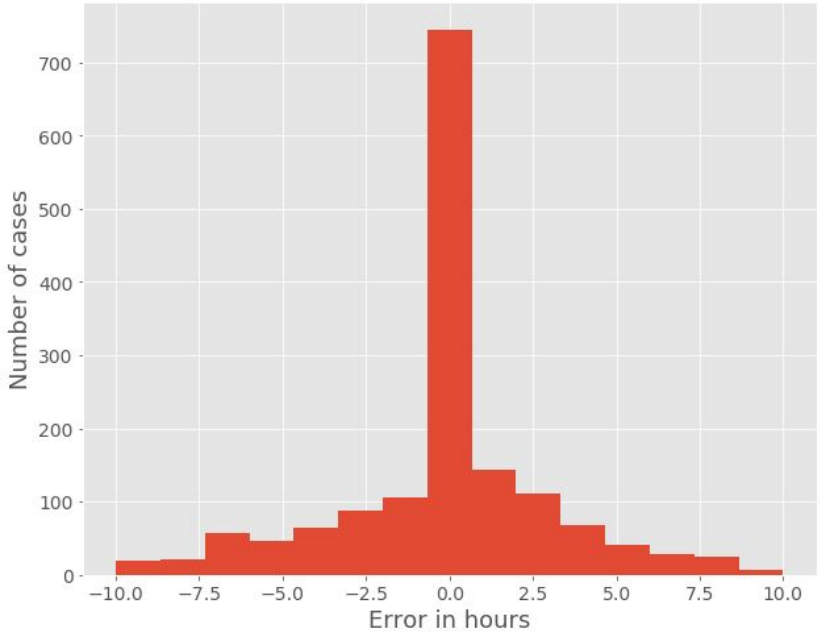
DataKind

Frequency distribution of errors

MVP Model MAE: 2hrs 24mins

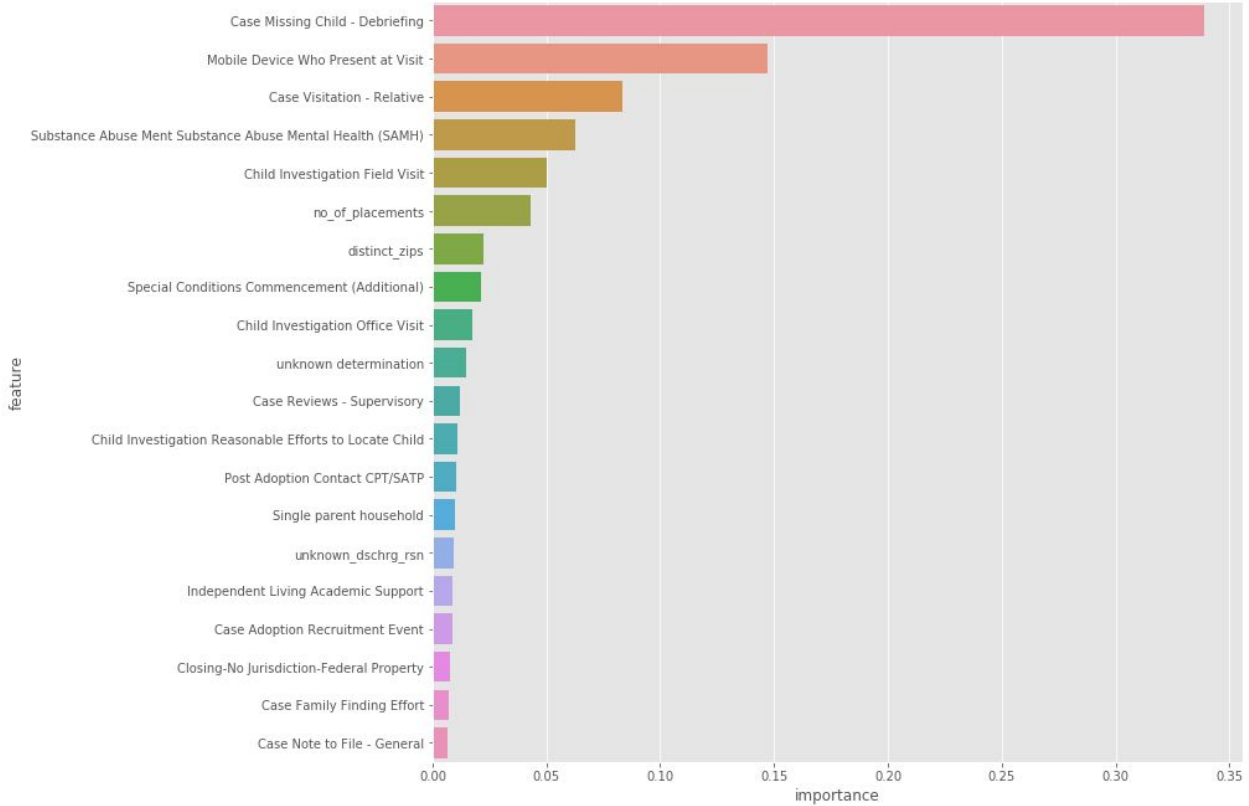


Current Model MAE: 4hrs 30mins



DataKind

Adding model interpretability



DataKind

Adding individual case-level interpretation

Case 5434

Predicted: 2.3 hours Actual: 2.0 hours

Most Significant Features:

1. Mobile Device Who Present at Visit
2. Case Missing Child - Debriefing
3. Number of zip codes
4. Case Visitation - Relative
5. Number of placements
6. Substance Abuse Ment Substance Abuse Mental He
7. Child Investigation Field Visit
8. Child Investigation Staffing - Multidisciplinary
9. Special Conditions Staffing - Red Flag
10. Case Family Finding Effort

Adding individual case-level interpretation

Case 49104

Predicted: 2.6 hours Actual: 14.5 hours

Most Significant Features:

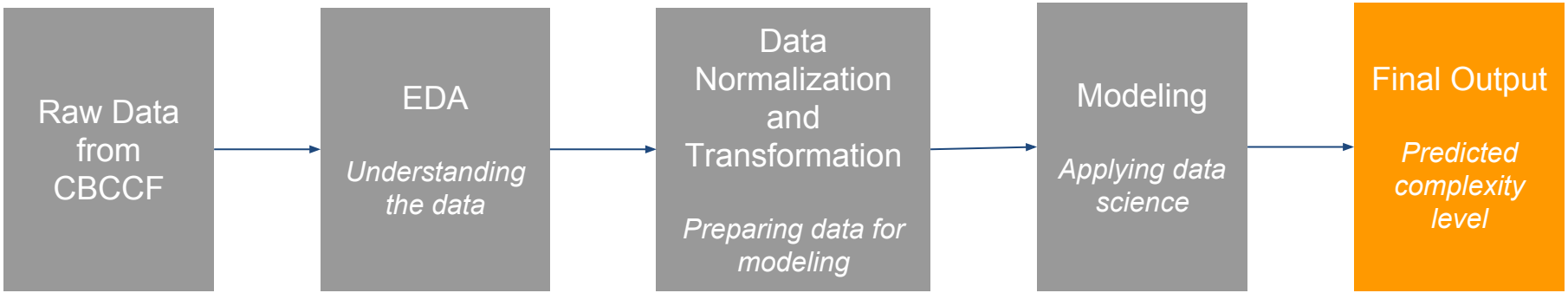
1. Case Visitation - Relative
2. Case Missing Child - Debriefing
3. Fl_alc_abus_prnt
4. Post Adoption Contact CPT/SATP
5. Mobile Device Who Present at Visit
6. Number of placements
7. Child Investigation Field Visit
8. Substance Abuse Ment Substance Abuse Mental
9. Number of zip codes
10. Closing-No Jurisdiction-Federal Property

DataKind

Major Error

- Some of the cases were being labelled with 0 activity

Our Process



DataKind

What We Are Predicting

What is the output of our model?

Our model offers a prediction of the level of complexity (low, medium or high) over the next 60 days for each case based on case attributes and activity from the previous 6 months.

- Low complexity = 0 - 4 hours per month
- Medium complexity = 5 - 16 hours per month
- High complexity = over 16 hours per month

Our model also attempts to shed some light on the factors that led to its predicted level of complexity.

Our Case Dataset (n = 9,280)

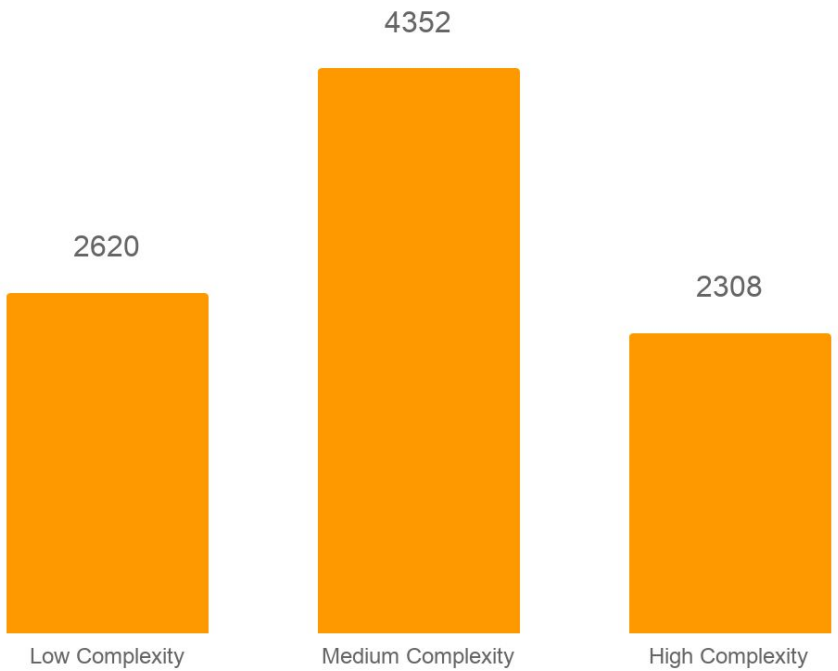
Eligible Cases:

- At least 3 months long but no longer than 18 years
- Case started after Jan 1, 2000
- Case has at least 5 notes

Case Activity:

- Mean: 14.5 hours
- Median: 7.5 hours
- Maximum: 380 hours

Number of Cases by Complexity Level



DataKind

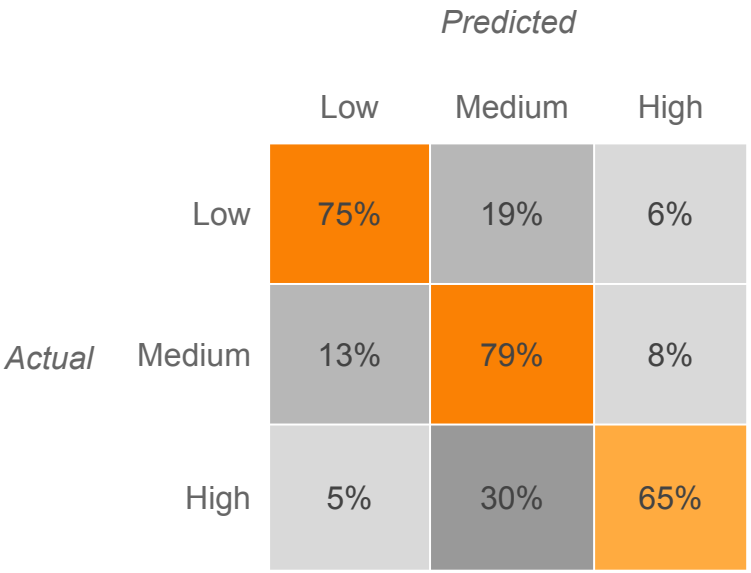
Our Model Features

Investigations: • Living arrangement • CPS type (in home or not) • Determination • Response priority	**Placements**: • Number of placements • Number of placement zip codes • Placement duration	**Removals**: • Manner of removal • Family structure • Discharge reason • Reason for service
Reason for Service: • Reasons for why child is considered at-risk	**Notes:** • Note category • Type of work being done	

DataKind

Our Prediction Accuracy

Confusion Matrix:

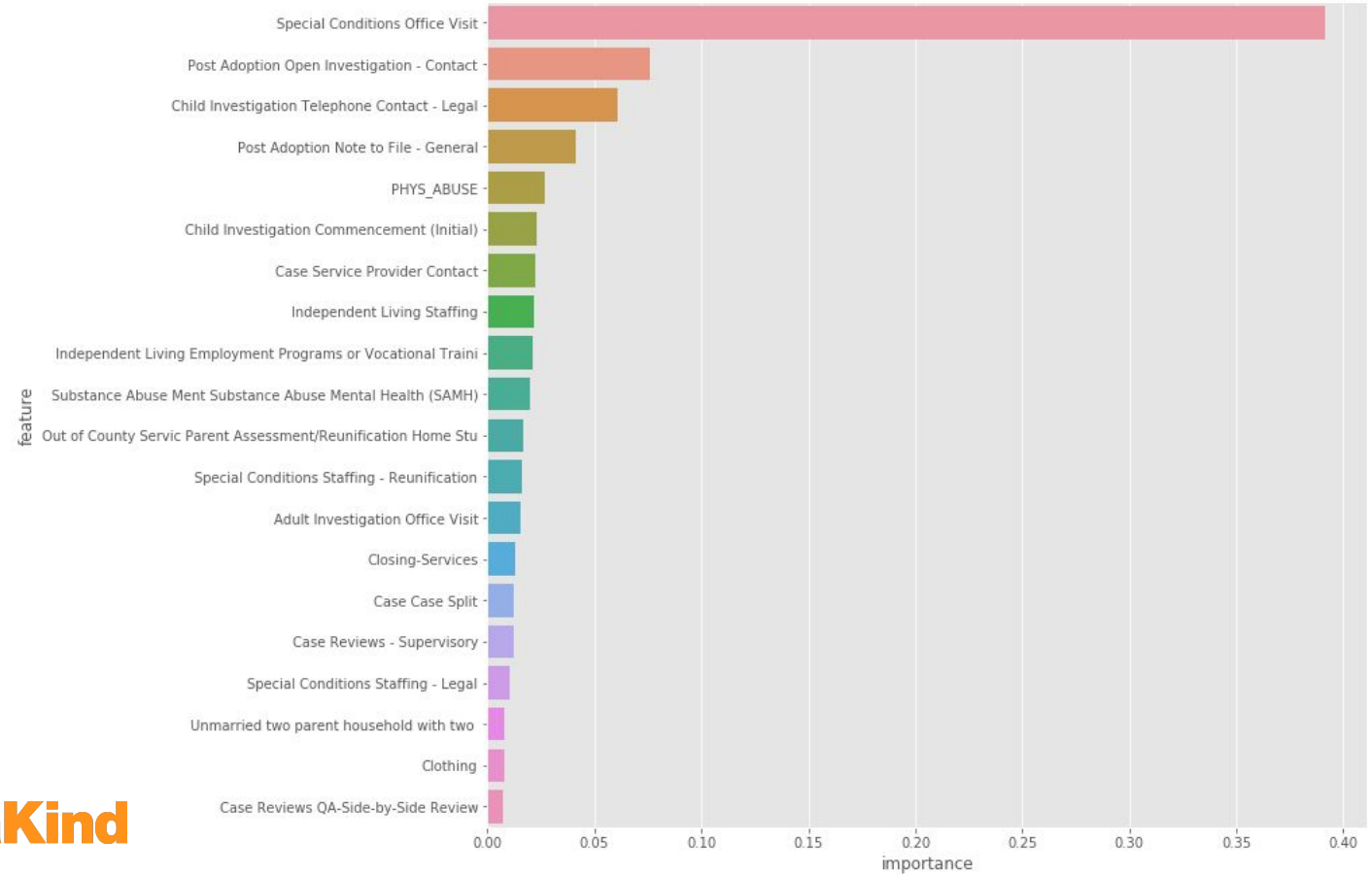


Overall, our model predicts the correct level of complexity 72% of the time (*precision*) and correctly classifies the complexity of 72% of all cases (*recall*).

	Precision	Recall
Low	0.69	0.71
Medium	0.71	0.79
High	0.78	0.59
Average	0.72	0.72

DataKind

The Most Important Features



DataKind

Example Use Case

- A supervisor wants to estimate the level of complexity for Case 9999797272 for the next 60 days
- Over the previous 6 months, Case 9999797272 had the following activities:
    - 16 Special Conditions Facsimile
    - 3 Independent Living Budget and Financial Management
    - 2 Adult Investigation Home Visit - VA’s Current Residence
    - 2 Child Investigation Office Visit
    - 2 Out of County Service Home Study
    - 2 Residential
    - 1 Special Conditions
    - 1 Case Home Study
    - And 9 other activities

Example Use Case

Model Prediction	Actual Hours Spent	Model Confidence
Medium Complexity	7	83%

The most important features identified for this case:

- Case Staffing - Reunification
- Clothing
- Special Conditions Commencement (Initial)
- Bed Hold
- Special Conditions Commencement (Additional)
- Special Conditions Staffing - Case Transfer/ESI
- Child Investigation Note to File - Interstate Compact
- Special Conditions Home Visit - Parent/Caregiver
- Case Visitation - Parent/CG Removed From
- Case Visitation - Sibling

Accomplished

- Complexity Model

Did not address:

- Caseworker specialty
- Geography (kind of)

Lessons Learned

- Modular Code (Faster Iterations)
- Partner with Strong IT background
  - Talk about deployment early
- Communicating Predictions
- Maybe should’ve considered the relationship of two projects more
- Experience with other stacks

Getting Involved in DataKind

- Apply Online
- Opening of Projects Varies

Volunteer Opportunity

The Problem

When a child is removed from their families, the goal is to place them in the best home first and for the child to remain in that same home for the duration of their time in foster care. Kids that move from foster home to foster home multiple times end up spending more time in foster care, have academic (many times school moves also occur) and attachment issues. See the Casey Family Foundation brief on Placement Stability for more information (https://www.casey.org/placement-stability-impacts/)

Currently, there are mechanisms in place (Placement Stability Staffings) to help a foster home when the placement is breaking down, but in many cases, by the time the foster parent requests a Placement Stability Staffing, it is too late.

Volunteer Opportunity

The Challenge

Using sentiment analysis on case notes (and possibly other data points), can we predict a placement disruption before the foster parent requests a change. If so, we can suggest a Placement Stability Staffing or other supports before the foster parent needs to ask.

We hope that this will improve stability and ultimately lead to better outcomes for the kids we care for.