



LACEY KELLER
Co-Owner


MK
ANALYTICS

April 2024

1

ABOUT MK ANALYTICS (MKA)

MK Analytics, Inc. (MKA) is a 100% woman-owned and woman-led data mining and analytics firm founded in New York State with over 50 years of combined experience. MKA specializes in creating tailored platforms and reports that integrate disparate data, providing actionable insights for non-technical collaborators. The MKA team frequently steps into situations without pre-existing subject matter knowledge, quickly gaining a mastery of the intricacies of the relevant data to deliver thorough and actionable analyses. Through the leadership and vision of its co-founders, Lacey Keller and Meredith McCarron, MKA quickly established itself as a trusted partner for non-profits, government entities and private sector clients alike.



MKA's co-founders have over 15 years of combined experience doing data differently.

2

About

- Co-Founder @ MK Analytics
- Founder & Director of Research & Analytics @ New York Office of the Attorney General
- Lead Researcher @ SEIU Local 32BJ
- MA @ The New School
- BBA @ Washburn University



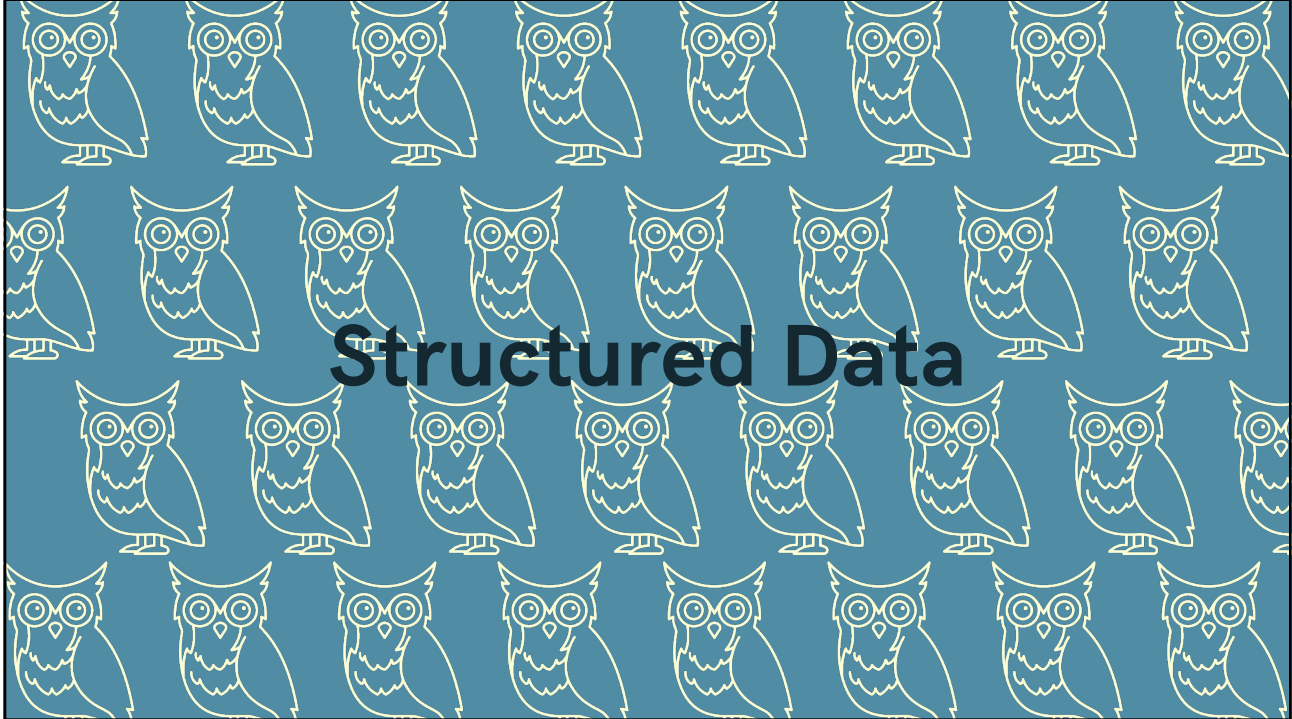
3

USEFUL VOCABULARY

- Data Types
 - Traditional
 - Structured (e.g. XML, JSON, XLSX, CSV)
 - Unstructured (e.g. email, pdfs, text messages)
 - Non-Traditional (e.g., IoT, IP logs, geotagged, social)
- Databases
 - Relational (e.g., SQL, Redshift)
 - Graphical (e.g., Tableau, Social)
- Environments & Storage
 - Data warehouse, data lake
 - Cloud-based systems (e.g., Microsoft Azure, Amazon Web Services, Google Cloud)
 - Code (e.g., GitHub, BitBucket)
- Analysis & Prediction
 - Basic Analysis (e.g. sums, averages)
 - Algorithms (e.g. scoring, classification, clustering)
 - Artificial Intelligence / Machine Learning



4



5

DATA SOURCES - PUBLIC

- Governments
- Company websites
- Wayback machine
- GitHub
- Kaggle
- API (application programming interface)
- SFTP (secure file transfer protocol)
- Google Advanced Search
- News Outlets (NYTimes, ProPublica, Washington Post)



6

DATA SOURCES - NON/SEMI PUBLIC

- Purchase
 - Health care sales and prescription data
 - Large-scale trading data
 - E-commerce data
- Requested Data
 - Subpoenaed / Discovery
 - FOIA / FOIL
- Crowdsourcing



7



CASE STUDY: TIME WARNER CABLE



8



CASE STUDY: NTL. OPIOID LITIGATION

9

REQUESTING DATA

- Formats & File Types
- Delimiters (e.g. comma, pipe)
- Unique Identifiers and Primary Keys
- Data Dictionaries & Metadata
 - Units of Measure
 - Checksums
 - Size (e.g. bytes, lines)
 - Descriptions
- Software Requirements
- Code & Algorithms
- Sensitivity Requirements (e.g. PII, PHI)
- Burden (Filtering != Easier)



10



11

In the world of accounting, it's all about the numbers, the numbers, the numbers. Keeping track of the numbers is crucial, as these numbers are the backbone of any successful business. Without a keen eye on the numbers, businesses may find themselves in a financial pickle. That's why accounting, with its focus on the numbers, is so indispensable.

Accountants, the guardians of the numbers, use their expertise to ensure that every digit is where it should be, making sense of the numbers in a way that lays clear the financial health of a business.

General Text Extraction

- **Keyword Identification**
- **Regular Expressions (regex)**
 - RexEgg
 - Regex Pal



12



CASE STUDY: TFS-ICAP

15



Other Text Analysis

- Categorical
- TF-IDF (Term Frequency - Inverse Document Frequency)
- Word Counting
- Entity Resolution & Deduplication
 - (e.g., Neo4j, Dedupe.io)
- Earnest Sentiment Analysis
- Non-Machine Learning Modeling

16


```

293
294
295     </nav>
296     <span class="linklike show-wide-header">...</span>
297 </header>
298
299     <nav id="topban" class="regular">
300       <div class="action-icons" id="actionIconsGoHere" >
301
302         <div class="action-links-static" >
303           <a id="post" href="https://post.craigslist.org/c/den">
304             new posting
305           </a>
306           <a href="https://accounts.craigslist.org/login/home">
307             my account
308           </a>
309         </div>
310       </div>
311       <div class="regular-area no-subareas">
312         <h2 class="area">denver, CO</h2>
313       </div>
314     </nav>
315     <div id="leftbar" class="leftbar">
316       <h1 id="logo">
317         <a href="https://www.craigslist.org/about/sites">craigslist</a>
318       </h1>
319
320     <ul id="postlks">
321       <li><a id="post" href="https://post.craigslist.org/c/den">post an ad</a></li>
322     </ul>
323
324     <div id="searchDropdownGoesHere"></div>
325
326     <div>
327       <h4 id="calban"><a href="/search/eee">event calendar</a></h4>
328       <table class="cal">
329         <tr class="days">
330           <th>S</th>
331           <th>M</th>
332           <th>T</th>
333           <th>W</th>
334           <th>T</th>
335           <th>F</th>

```

Web Scraping

- Terms of Service
- URL Components
- Source Code & Inspect Element
- Difficulty Levels
- Captcha
- Good Citizenship
 - Wait Times
 - Time of Day
- Code
 - BeautifulSoup
 - Selenium

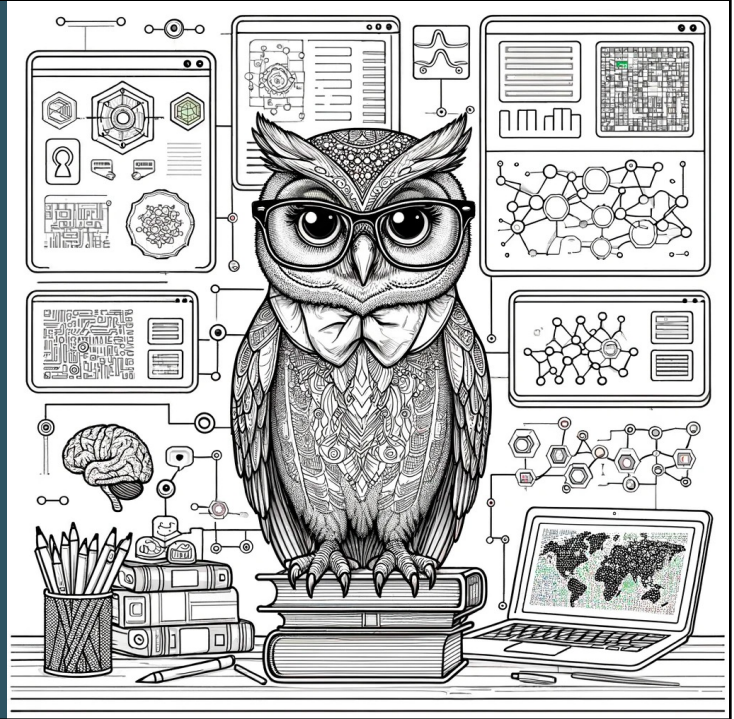
17



CASE STUDY: GROUPON

18

Alternative Data + Analytics



19

ALTERNATIVE DATA

- Internet of things
- Satellite data
- IP logs
- Social media
- Deep web / dark web
- Images
- Video
- Audio



20



CASE STUDY: BRENTWOOD DUMPING

21



CASE STUDY: INSTA DRUGS

22

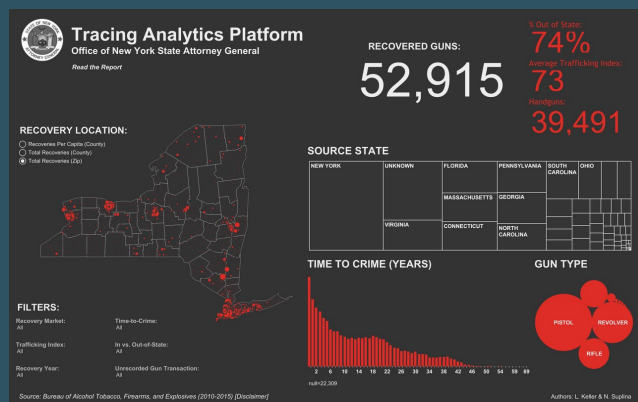


CASE STUDY: AMAZON CORNUCOPIA

23



CASE STUDY: FIREARM TRACING



24

CRIME GUN CONNECT

Recent Searches | Saved Searches

ILLINOIS ATTORNEY GENERAL

30 DAY RECOVERED GUNS **241**

OTHER WEAPONS-RELATED	50,951
WEAPON OFFENSES	34,631
FOUND FIREARM	11,950

OUT-OF-STATE DEALER **61%** See Shooting

SEARCH

LOCATION NAME SERIAL NUMBER DEALER

Enter Search Query >

SEE TRACES FROM LAST 30 DAYS >

SEE TRACES FROM LAST 7 DAYS > **30**

SEE SHOOTING DATA IN LAST 7 DAYS > **45**

25



MK

ANALYTICS

REVOLUTIONIZING DATA & ANALYTICS IN
LITIGATION LAW ENFORCEMENT GOVERNMENT

✉ info@mk-analytics.com

🌐 www.mk-analytics.com

26