# SCALING REVENUE THROUGH DATA SCIENCE

Joseph Roberts

Matt Rubino

Che Guan

https://github.com/mids-w205/project-3-joethequant

Acme Gourmet Meals

Acme Gourmet Meals (AGM) is an innovative food company that offers healthy, seasonal pre-made meals. It delivers to customers in the Greater Bay Area, focusing on young professionals that are too busy to cook or grocery shop.

# 01 BACKGROUND

DATA SCIENCE

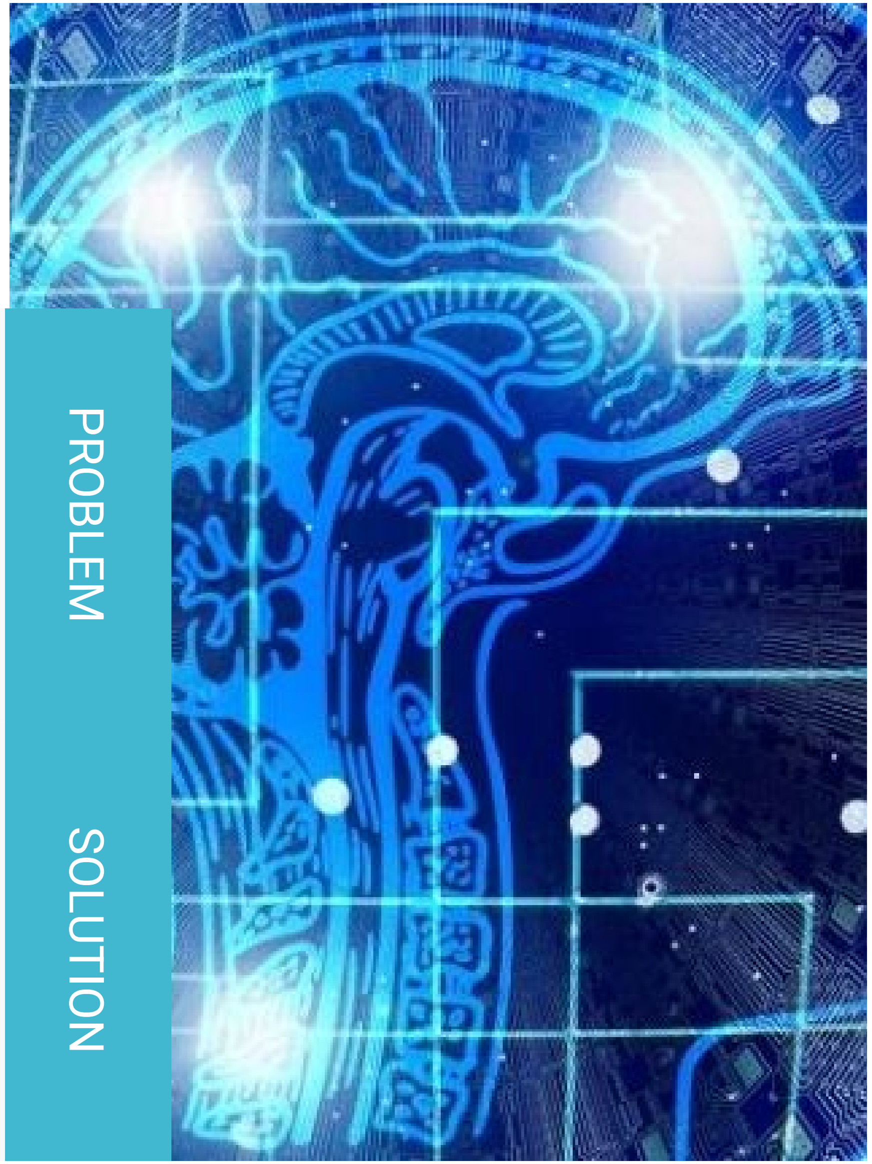
# A VISION FOR THE FUTURE

PROBLEM

**Problem Statement:** How can AGM optimize delivery and pickup capabilities in the Bay Area to better serve their customers today while positioning the company for long-term growth and success?

SOLUTION

**Continuous Learning through Data Science:** The innovative management team is focused on expanding its store footprint, optimizing delivery in the Bay Area, and iterating on its successes with help of the data science team to drive new ideas, insights, and intuition. It believes the use of data science can lead to an improved margin profile, increased TAMs, and take rates.



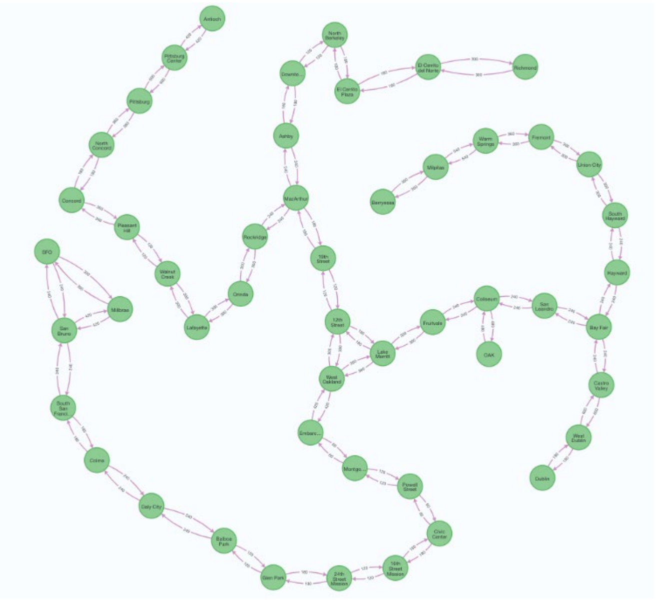
# Questions of Interest

- What is our target market from the perspective of customer demographics? What are their ages, genders, income levels and primary places of residence?
- What times during the day and week tend to lead to peak sales generation? What can we learn from the surrounding areas at such times?
- What features will be the best predictor for increasing sales?
- Does Betweenness make sense as a variable to optimize?
- How will we test and roll out new ideas?
- Can we optimize our inventory more efficiently?

# Features

- Degree Centrality
  - Measures number of relationships a node has in a graph: incoming, outgoing
- Closeness Centrality
  - Measures average of the shortest path distances between a node and all other nodes
- Betweenness Centrality
  - Find all pairs shortest paths (weighted)
- Triangle Count
  - Number of triangles that pass through a node
- Clustering Coefficient
  - Probability that neighbors of a node are connected to each other
- Louvain Modularity
  - Creates a hierarchy of group at different scales
- Population within X miles of a zip code
- Number of Stations within X miles of a zip code

Visualize BART stations and travel time between any two



| | node | degree | closeness | betweenness | triangle_count | clustering_coefficient | community |
|---|---|---|---|---|---|---|---|
| **0** | 12th Street | 3 | 0.2333 | 1116 | 1 | 0.3333 | 29 |
| **1** | 16th Street Mission | 2 | 0.1712 | 720 | 0 | 0.0000 | 10 |
| **2** | 19th Street | 2 | 0.2195 | 1088 | 0 | 0.0000 | 46 |
| **3** | 24th Street Mission | 2 | 0.1663 | 656 | 0 | 0.0000 | 10 |
| **4** | Antioch | 1 | 0.1069 | 0 | 0 | 0.0000 | 31 |
| **5** | Ashby | 2 | 0.1986 | 440 | 0 | 0.0000 | 2 |

| zip | latitude | longitude | city | state | population | area | density | time_zone | within_1m_station | within_2m_station | within_3m_station |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 90011 | 34.0071 | -118.2587 | Los Angeles | CA | 109414 | 4.5137 | 24240.49 | America/Los_Angeles | 0.0 | 0.0 | 0.0 |
| 90650 | 33.9069 | -118.0826 | Norwalk | CA | 105886 | 9.8416 | 10759.07 | America/Los_Angeles | 0.0 | 0.0 | 0.0 |
| 91331 | 34.2556 | -118.4208 | Pacoima | CA | 105799 | 9.0328 | 11712.71 | America/Los_Angeles | 0.0 | 0.0 | 0.0 |
| 90201 | 33.9707 | -118.1708 | Bell Gardens | CA | 102433 | 6.0050 | 17057.93 | America/Los_Angeles | 0.0 | 0.0 | 0.0 |
| 92335 | 34.0872 | -117.4655 | Fontana | CA | 99284 | 18.0598 | 5497.51 | America/Los_Angeles | 0.0 | 0.0 | 0.0 |

# STRATEGY IMPLEMENTATION TIMELINE

01. BART

02. Additional Delivery Systems

03. Store Footprint Expansion & Iterate

04. Continuous Supply Chain Optimization Through Live Databases

### BART

Target Bart Stations that have a high population and high betweenness with other stations.

### Additional Delivery Systems

Drones can be used from remote delivery in suburban areas while robots can be used for more urban settings that do not have immediate Bart locations

### STORE FOOTPRINT EXPANSION & ITERATE

Targeting densely populated areas with no access to stores or trains currently

### CONTINUOUS SUPPLY CHAIN OPTIMIZATION

Leverage software stacks including Neo4j, MongoDB, Redis and SQL to optimize the supply chain on a live basis

# PHASE 1: BART

## 01. POWELL STREET

(i) High population with 140k people located within one mile.
(ii) Central connector between other stations.
(iii) Added benefit of being located in business district.

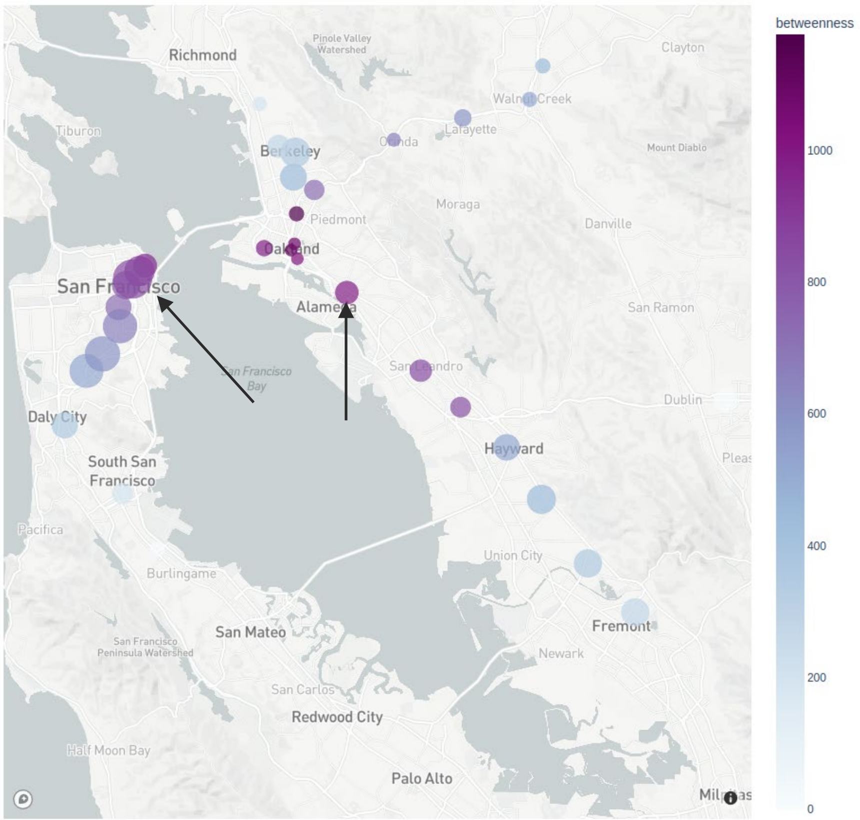
## 02. FRUITVALE STATION

(i) High population with 52k people located within one mile.
(ii) Central connector between other stations.

### TARGET

Bart Stations that have a high population and high betweenness with other stations.

Station Betweenness and Population within 1 Mile

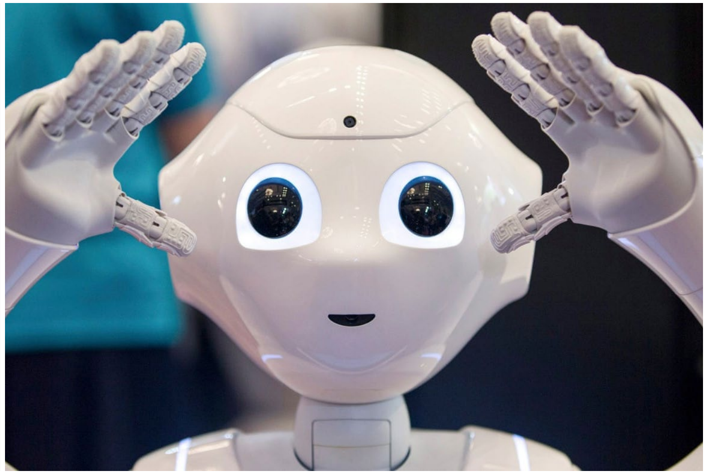


Drones & Robots

# PHASE 2: ADDITIONAL DELIVERY SYSTEMS

EFFICIENT DELIVERY

(i) To cater to the Non-Bart population, we propose using drones and delivery bots so as to maximize efficiencies and speed time to delivery
    (i) Drones can be used from remote delivery in suburban areas while robots can be used for more urban settings that do not have immediate Bart locations



# PHASE 2: COST OF ALTERNATIVE DELIVERY SYSTEMS

**DRONES**

(i) We estimate each drone delivery to cost around $5 per package

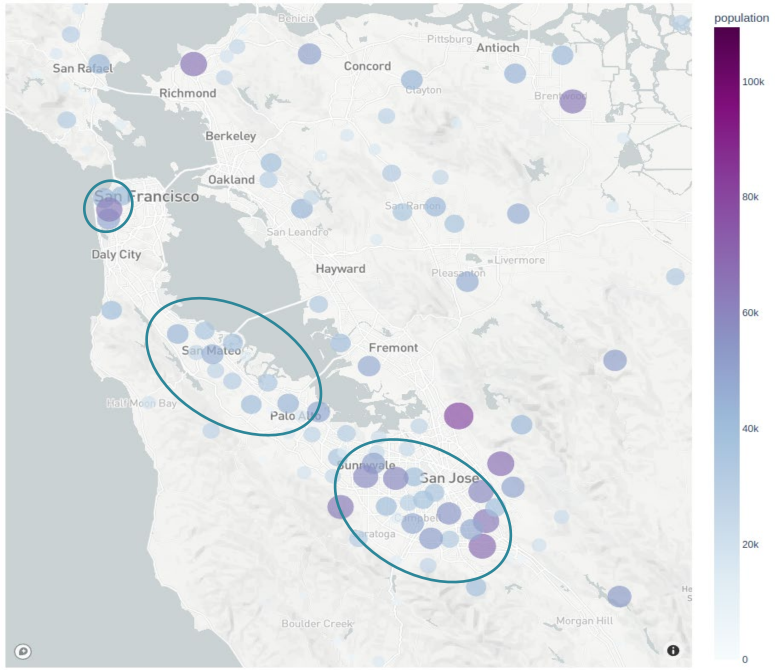
(ii) To begin building and testing a fleet, we recommend buying 3 camera drones, which cost about $400 per drone

**ROBOTS**

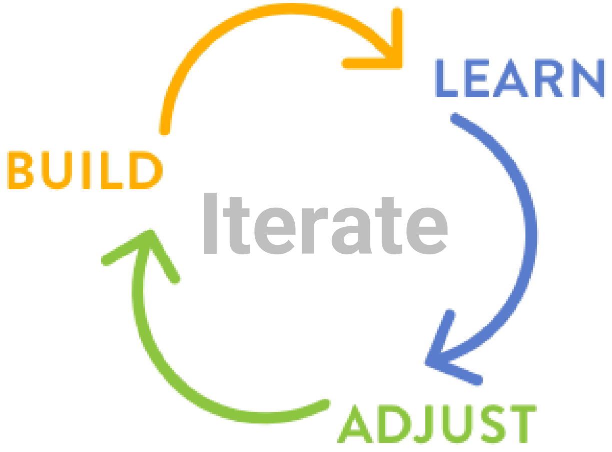
(i) We recommend purchasing a mix of both long range and short range robots, which we project to incur costs of $2.5k to $5k per robot

# PHASE 3: STORE FOOTPRINT EXPANSION & ITERATE

Population of Zip Codes, Station Greater Than 2 Miles Away




- Two-pronged approach to store footprint expansion:
  - Seeking densely populated areas with no train access and limited store alternatives




**San Jose, CA**

**San Mateo, CA**

**San Francisco, CA**

### INVENTORY TRACKING

Reduce manual error, minimize order cancellations, and identify inventory shrinkage

### INVENTORY OPTIMIZATION

Extending the working capital and cash conversion cycle through forecasting


# PHASE 4: CONTINUOUS SUPPLY CHAIN OPTIMIZATION THROUGH LIVE DATABASES

### ROUTE OPTIMIZATION

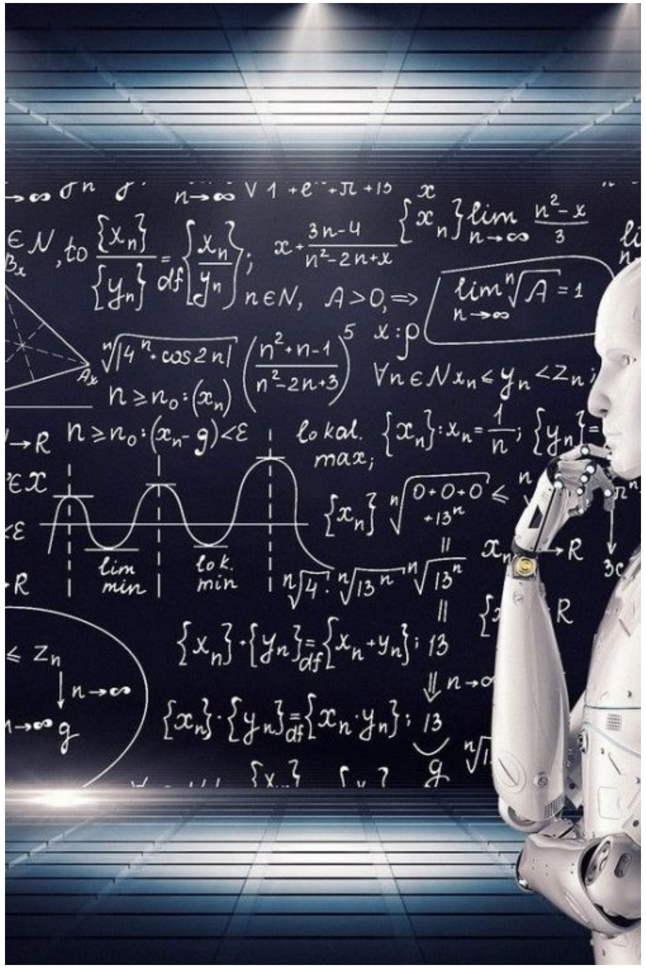
Modelling for shortest delivery times and minimal interference

### LIVE DELIVERY TRACKING

The logistics team and customers can be informed on live deliveries through our enhance delivery systems (drones and bots)

# LEVERAGING TECHNOLOGIES

Leveraging a full suite of technology stacks to drive innovation



neo4j

Measuring connections and characteristics of stores, BART stations and alternative delivery systems

SQL

Highly normalized transactional database for all operations.

mongoDB

Analytical dashboards to drive business viewpoints, capital allocation, and mix shifts

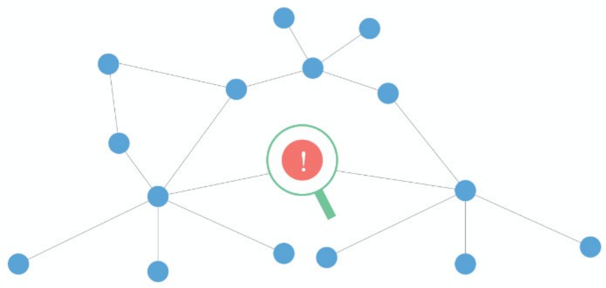
redis

Real time date based on inventory, wait times, and sales

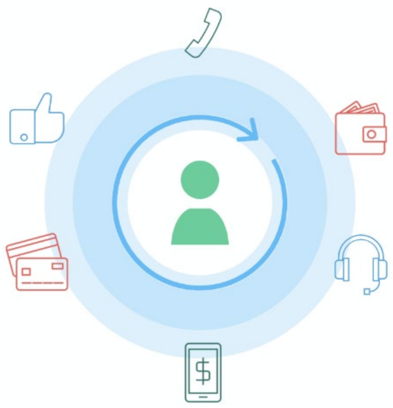
- Relational databases aren’t designed to capture rich relationship information. Applications are missing out on critical connections essential for today’s high-stakes, data-driven decisions.
- Neo4j graph data science makes it easier to unlock answers as it puts relationships first.  It also allows one to use graph queries and algorithms to further investigate data and prepare features for ML models to use.

Graph Technology in Financial Services

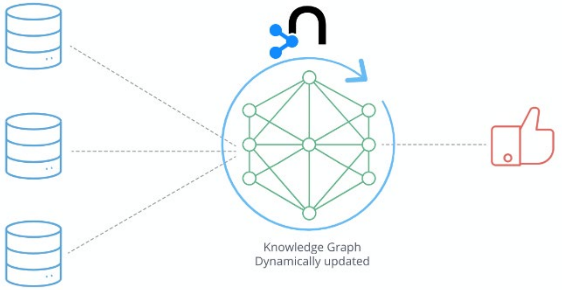
Stay Agile in Risk & Compliance



Fraud Detection

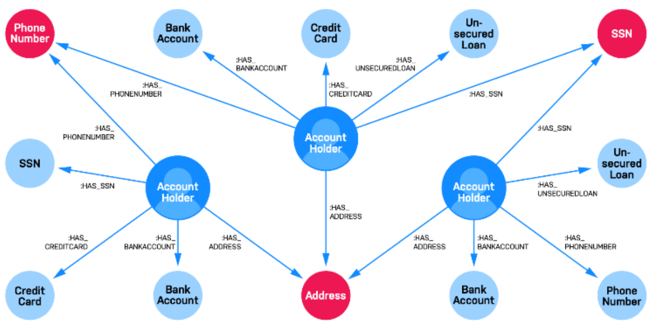


Capture a 360 Customer View



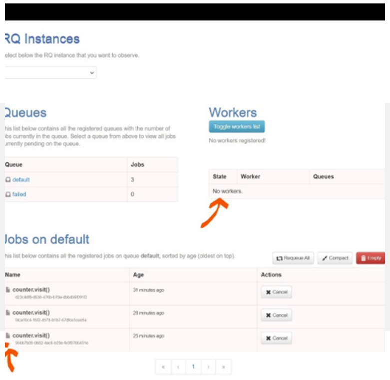

Leverage Data across Team

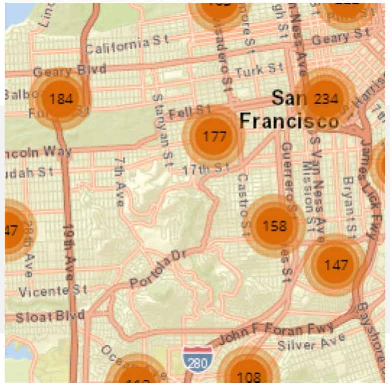
A Business Use Case: Identifying First-Party Fraud




- An individual or group of people may misrepresent their identity or use false information when applying for a financial product or service. An estimated 80% of all credit card fraud losses stem from this synthetic identity fraud which results in major losses for financial institutions.
- To detect such first-party fraud efficiently and effectively, Neo4j allows one to use graph queries and algorithms to investigate data, get a sense of its structure and discover patterns and anomalies.
- One can use Neo4j Bloom to visually explore results and verify first-party fraud. A more advanced approach is to apply neo4j graph algorithm to create features and use them to machine learning model to identify more fraud faster.

Redis is an open source, in-memory data structure server which leverages key-value pairs. The Redis stack permits the real time building of applications with searchable JSON, time series and graph data models, as well as high performance probabilistic data structures.



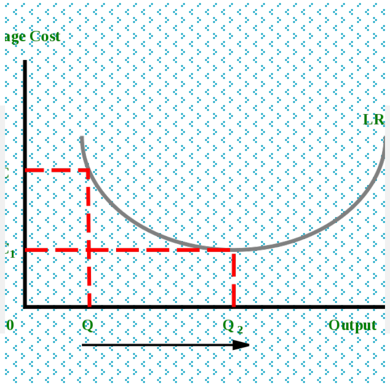




## LIVE USAGE DASHBOARD

Showing performance, memory, activity metrics

## OPTIMAL SCALE OUT

Sequentially copying of the file system of the master container into the newly created nodes


## SECURE USER LOGINS

Protecting against cyber-attacks with no way for others to reverse passwords

## INFORMATION ACCESS

Once logged in, users can access a full list of inventory, delivery pipeline and routes

## USE CASE

(i) Scaling out a NoSQL database that holds username and passwords for secure log-ins to protect against cyber-attacks

(ii) Post log-in, users would have the ability to see a full product list, delivery routes, and other alternative data

(iii) A stateful web server because it can track and hold session data for every transaction, allowing for optimal scale out mongoDB

Mango is an open source, data structure server which leverages a json like structure to store and retrieve data. The MangoDB creates a highly efficient way to access specific analytical views of our data.

Example: Store and Retrieve all data about each of our customers.

```
{
  "_id": "5cf0029caff5056591b0ce7d",
  "firstname": "Jane",
  "lastname": "Wu",
  "address": {
    "street": "1 Circle Rd",
    "city": "Los Angeles",
    "state": "CA",
    "zip": "90404"
  },
  "hobbies": ["surfing", "coding"]
}
```

## Customers

Name
Address
Email
Payment Details
...

## Stores

Location
Employees
Inventory
Sales
...

## Transactions

Customer
Sales Employee
Store
Order Details
...

## Analytics

Sales Trends
Inventory Mgmt.
Predictions
Customer Scores
...

mongoDB

## Applications

Online and Stores
Executive Dashboards
Research Datastore
Engagement Surveys
Sales Forecasting
Employee Specific
...

# FINANCIAL PROJECTIONS

- An investment in the data science team represents a significant opportunity from a business perspective
- Driving increased sales growth, longer-run operating margins in excess of 40% , and the ability to scale efficiently
- Perception of AGM as a technology-enabled company could lead to multiple re-rating

| | 2023 | 2024 | 2025 | 2026 | 2027 | 2028 | 2029 | 2030 | 2031 | 2032 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Phase 1 Revenue Model** | | | | | | | | | | |
| Local Population Exposure | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 | 192,500 |
| % Daily Sales | 0.4% | 0.6% | 0.9% | 1.4% | 2.0% | 2.0% | 2.0% | 2.0% | 2.0% | 2.0% |
| Orders | 281,050 | 421,575 | 632,363 | 948,544 | 1,422,816 | 1,422,816 | 1,422,816 | 1,422,816 | 1,422,816 | 1,422,816 |
| Average Daily Orders | 770 | 1,155 | 1,733 | 2,599 | 3,898 | 3,898 | 3,898 | 3,898 | 3,898 | 3,898 |
| Average Order Size | $15 | $15 | $15 | $15 | $15 | $15 | $15 | $15 | $15 | $15 |
| **Phase 1 Sales** | **4,215,750** | **6,323,625** | **9,485,438** | **14,228,156** | **21,342,234** | **21,342,234** | **21,342,234** | **21,342,234** | **21,342,234** | **21,342,234** |
| **Phase 2 Revenue Model** | | | | | | | | | | |
| Local Population Exposure | | | | 800,000 | 800,000 | 800,000 | 800,000 | 800,000 | 800,000 | 800,000 |
| % Daily Sales | | | | 0.4% | 0.6% | 0.9% | 1.4% | 2.0% | 2.0% | 2.0% |
| Orders | | | | 1,168,000 | 1,752,000 | 2,628,000 | 3,942,000 | 5,913,000 | 5,913,000 | 5,913,000 |
| Average Daily Orders | | | | 3,200 | 4,800 | 7,200 | 10,800 | 16,200 | 16,200 | 16,200 |
| Average Order Size | | | | $64 | $64 | $64 | $64 | $64 | $64 | $64 |
| **Phase 2 Sales** | | | | **75,008,960** | **112,513,440** | **168,770,160** | **253,155,240** | **379,732,860** | **379,732,860** | **379,732,860** |
| **Phase 3 Revenue Model** | | | | | | | | | | |
| Local Population Exposure | | | | | | 800,000 | 800,000 | 800,000 | 800,000 | 800,000 |
| % to Sales | | | | | | 0.4% | 0.6% | 0.9% | 1.4% | 2.0% |
| Number of Sales | | | | | | 1,168,000 | 1,752,000 | 2,628,000 | 3,942,000 | 5,913,000 |
| Average Order Size | | | | | | $64 | $64 | $64 | $64 | $64 |
| **Phase 3 Sales** | | | | | | **74,752,000** | **112,128,000** | **168,192,000** | **252,288,000** | **378,432,000** |
| Total Orders | 281,050 | 421,575 | 632,363 | 2,116,544 | 3,174,816 | 5,218,816 | 7,116,816 | 9,963,816 | 11,277,816 | 13,248,816 |
| **Total Sales** | **4,215,750** | **6,323,625** | **9,485,438** | **89,237,116** | **133,855,674** | **264,864,394** | **386,625,474** | **569,267,094** | **653,363,094** | **779,507,094** |
| **Cost of Revenue** | **2023** | **2024** | **2025** | **2026** | **2027** | **2028** | **2029** | **2030** | **2031** | **2032** |
| Number Bart Locations | 3 | 3 | 3 | 3 | 3 | 6 | 6 | 6 | 6 | 6 |
| Bart Rent (locations *12 * $2000) | (72,000) | (72,000) | (72,000) | (72,000) | (72,000) | (144,000) | (144,000) | (144,000) | (144,000) | (144,000) |
| Drones and Robots Cost (Orders * $5) | - | - | - | (5,840,000) | (8,760,000) | (13,140,000) | (19,710,000) | (29,565,000) | (29,565,000) | (29,565,000) |
| Phase 3 Locations | | | | | | 3 | 3 | 3 | 3 | 3 |
| Phase 3 Rents (Locations * 12 * $5000) | - | - | - | - | - | (180,000) | (180,000) | (180,000) | (180,000) | (180,000) |
| Cost of Goods Sold (Sales * 20%) | (843,150) | (1,264,725) | (1,897,088) | (17,847,423) | (26,771,135) | (52,972,879) | (77,325,095) | (113,853,419) | (130,672,619) | (155,901,419) |
| Other Costs | (1,264,725) | (1,897,088) | (2,845,631) | (26,771,135) | (40,156,702) | (79,459,318) | (115,987,642) | (170,780,128) | (196,008,928) | (233,852,128) |
| Server Budget | (60,000) | (72,000) | (86,400) | (103,680) | (124,416) | (149,299) | (179,159) | (214,991) | (257,989) | (309,587) |
| # DS Employees | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 12 | 14 | 17 |
| Data Science Team ( $500,000 annual) | (1,500,000) | (2,000,000) | (2,500,000) | (3,000,000) | (3,500,000) | (4,000,000) | (5,000,000) | (6,000,000) | (7,000,000) | (8,500,000) |
| **Net Revenue** | **2023** | **2024** | **2025** | **2026** | **2027** | **2028** | **2029** | **2030** | **2031** | **2032** |
| Total Operating Costs | (3,739,875) | (5,305,813) | (7,401,119) | (53,634,238) | (79,384,253) | (150,045,496) | (218,525,896) | (320,737,538) | (363,828,536) | (428,452,134) |
| **Net Revenue** | **475,875** | **1,017,813** | **2,084,319** | **35,602,878** | **54,471,421** | **114,818,898** | **168,099,578** | **248,529,556** | **289,534,558** | **351,054,960** |
| **Operating Margin** | **11%** | **16%** | **22%** | **40%** | **41%** | **43%** | **43%** | **44%** | **44%** | **45%** |

# Questions?