Jumpman23
Market and Data Integrity Issues

01 **Data Overview & Data Integrity**

02 **Fraud and risk in the Market**

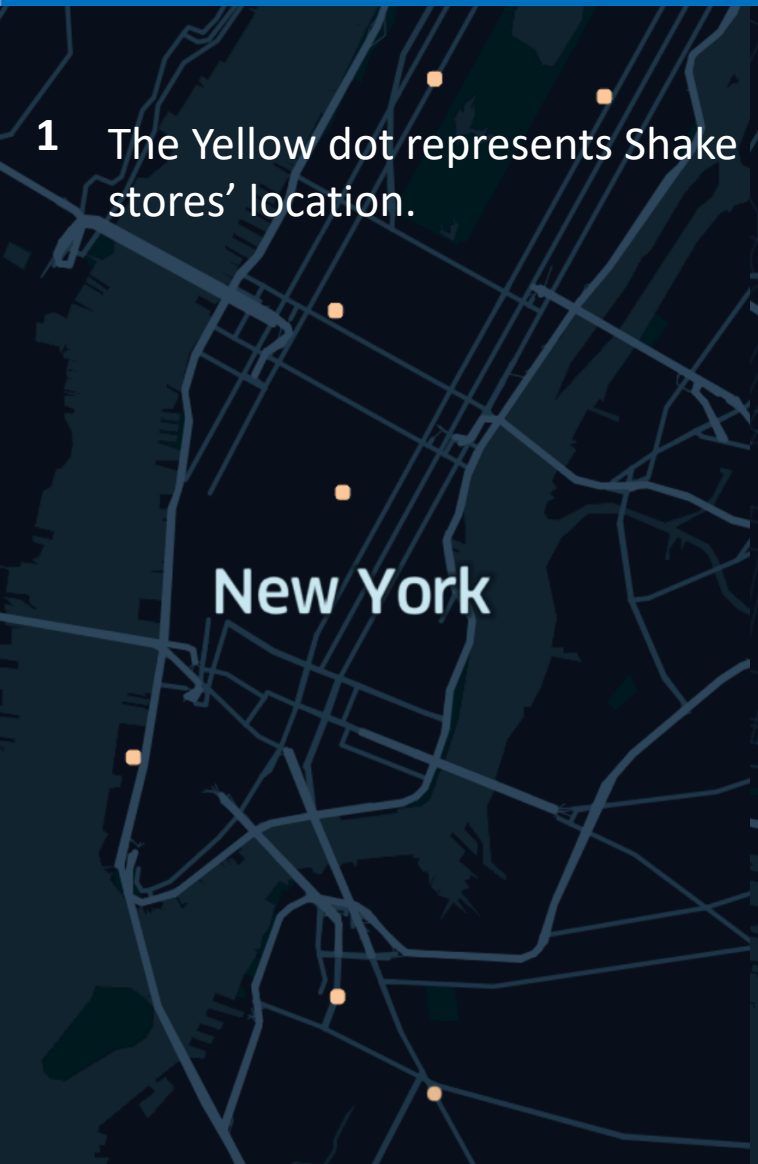
Postmates

Data Overview - Values

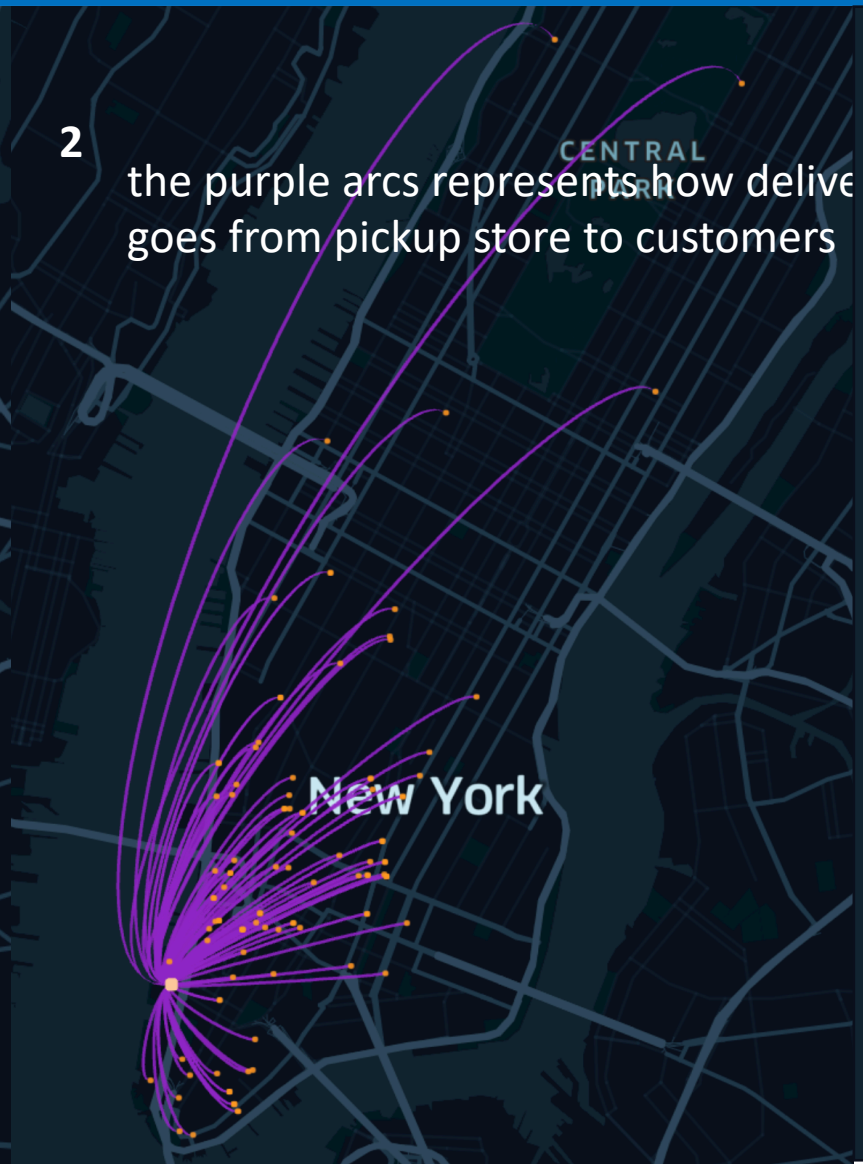
Data Types	Examples
Unique Identifiers	Delivery ID, jumpman ID, customer ID
Numerical	Order Quantity
Categorical	Product Type, name, pick up place
Temporal	Start time, end, arrival, drop off
Geological	latitude, longitude

We took 6K orders in Oct 2014, from 3.2K users, completed by 600 jumpmen, from 900 merchants

Case --- Shake Shack (restaurant with the most orders)



1 The Yellow dot represents Shake stores' location.



2 the purple arcs represents how delive goes from pickup store to customers

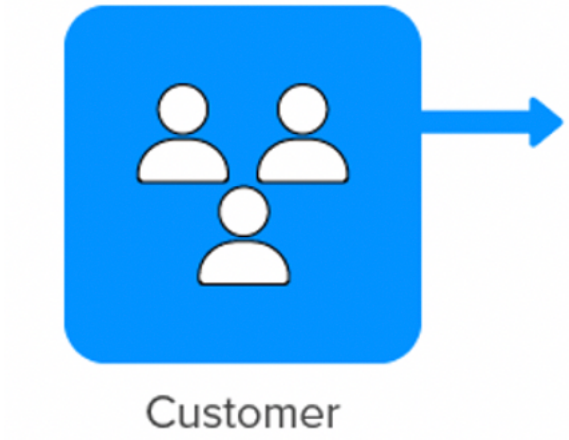
3 Jumpman's customers X pick the closet store.

Assumption:
products among Shake Shack have no difference.

4 Recommended action:

- use a distance measure to match stores & customers:
- Advice customer to enter correct zipcode
- Update in-app/ website search results by closest store set as default

Data Overview - Facts



Geo Segments:

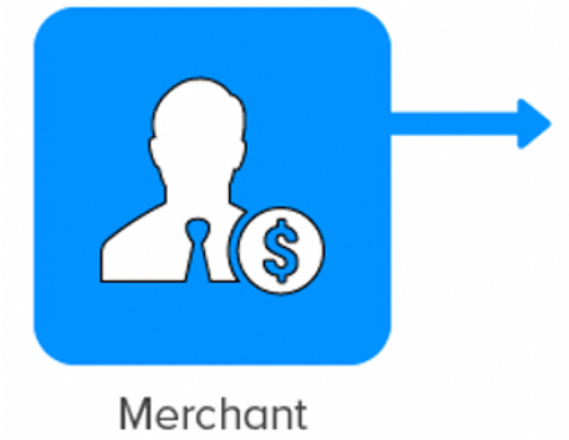
96% in Manhattan
4% in Brooklyn

Frequent Order time:

12pm, 7 pm
Weekends

Repurchase (week) rate:

30%



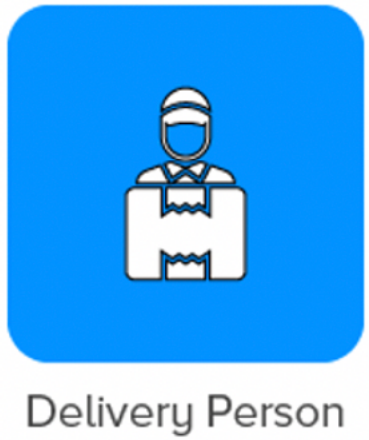
Preparation time

25%: 22 minutes
Mean: 31 minutes
75% : 43 minutes

30% merchants have average preparation time over 50 minutes.

Item Categories:

Food category: 57
total Item: 2277



Delivery time:

25%: 33 minutes
Mean: 43 minutes
75%: 55 minutes

47.5% of delivery person make less than 5 orders in a month

22.9 % of orders missing in store check- in

Market Analysis

40 -> 1st week in Oct 2014
44 -> last week in Oct 2014

Jumpman

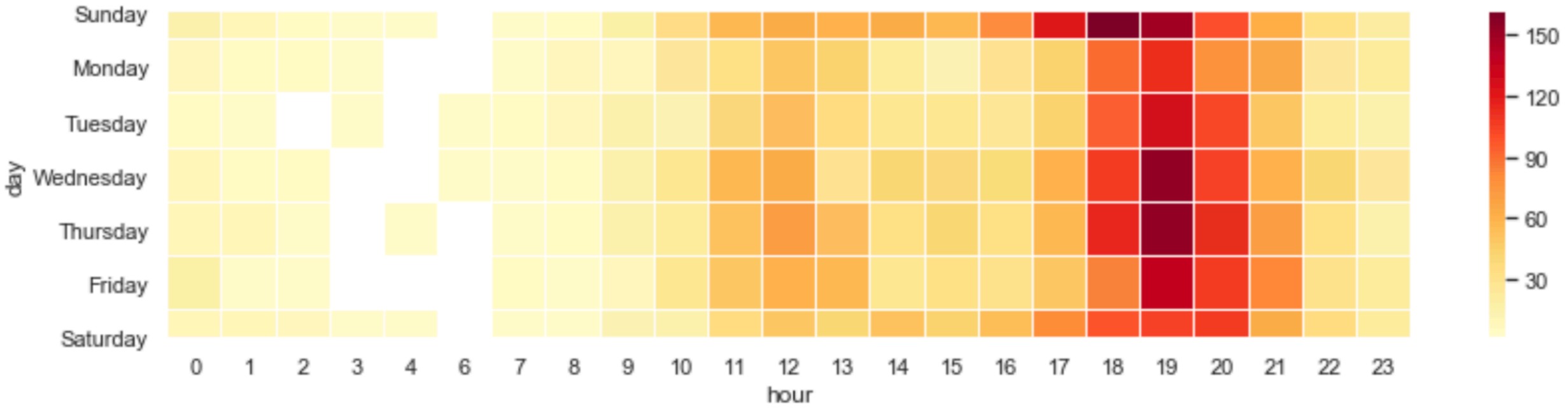
- **578** Jumpmen take orders in Oct 2014,
- **47.5%** of the Jumpmen has less than make 5 orders

Merchants

- Average preparation time **18 minutes**
- **30%** (268/898) merchants have average preparation time over 50 minutes.

Customers

- They place more orders at weekends (especially on Sunday) than during weekdays.
- The order peaks are at noon (12 pm) and in the evening (7 pm).
- Around 70% of the customers only order once in 30 days.

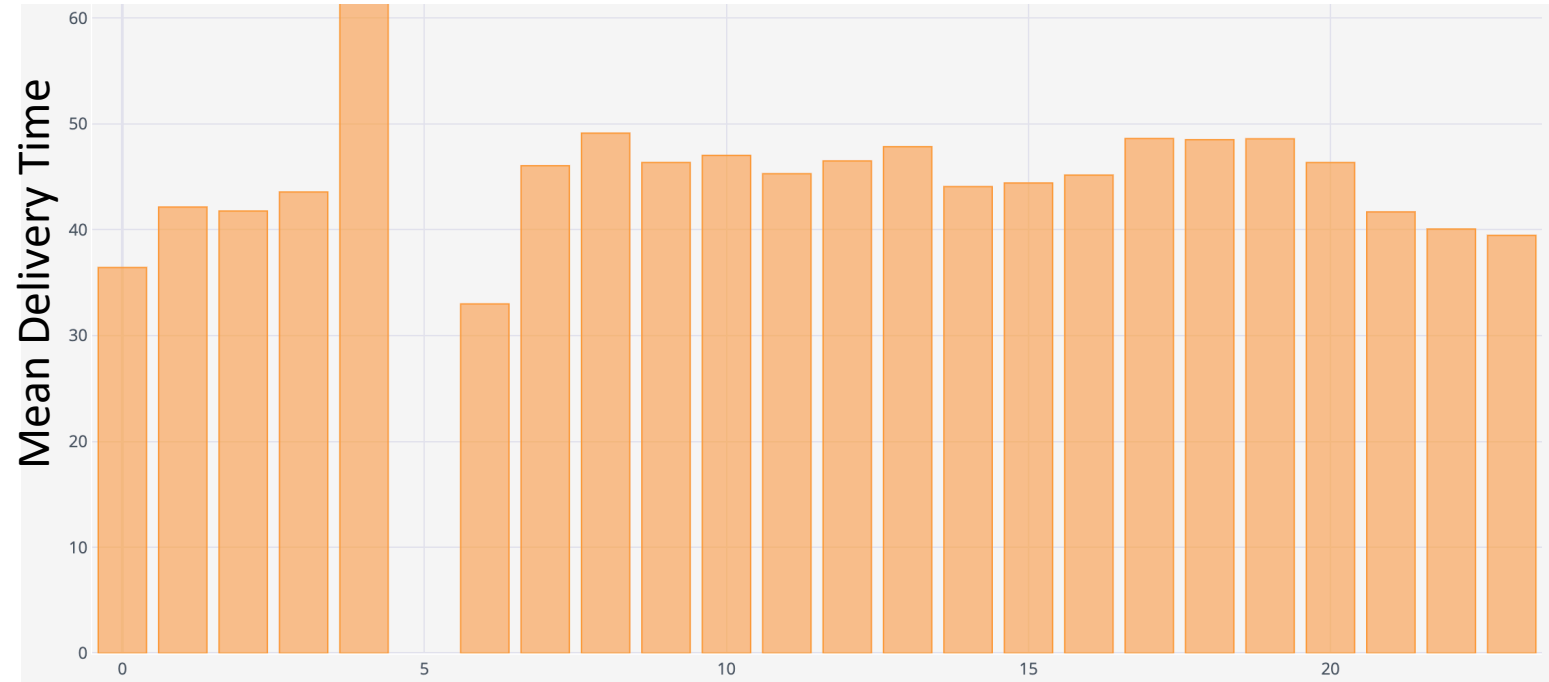


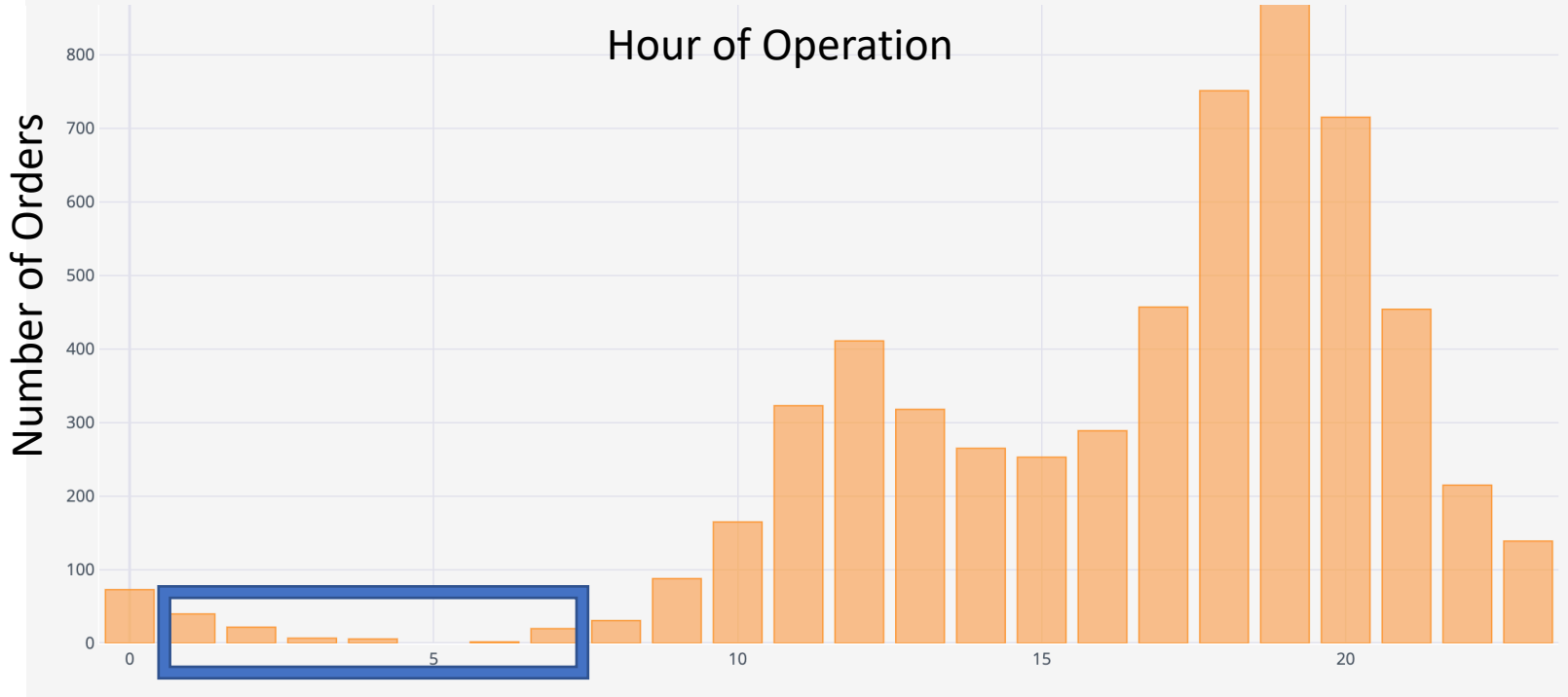
Delivery

Turns out, in Oct 2014, all the orders during “off_peak” hour represents **1.2%** of total orders.
(72 / 5983 records)

Not profitable if our business won't growth tremendously.

Off_Peak hour: 2am~ 6am





Data Integrity due --- System Error

3 Common Value issues

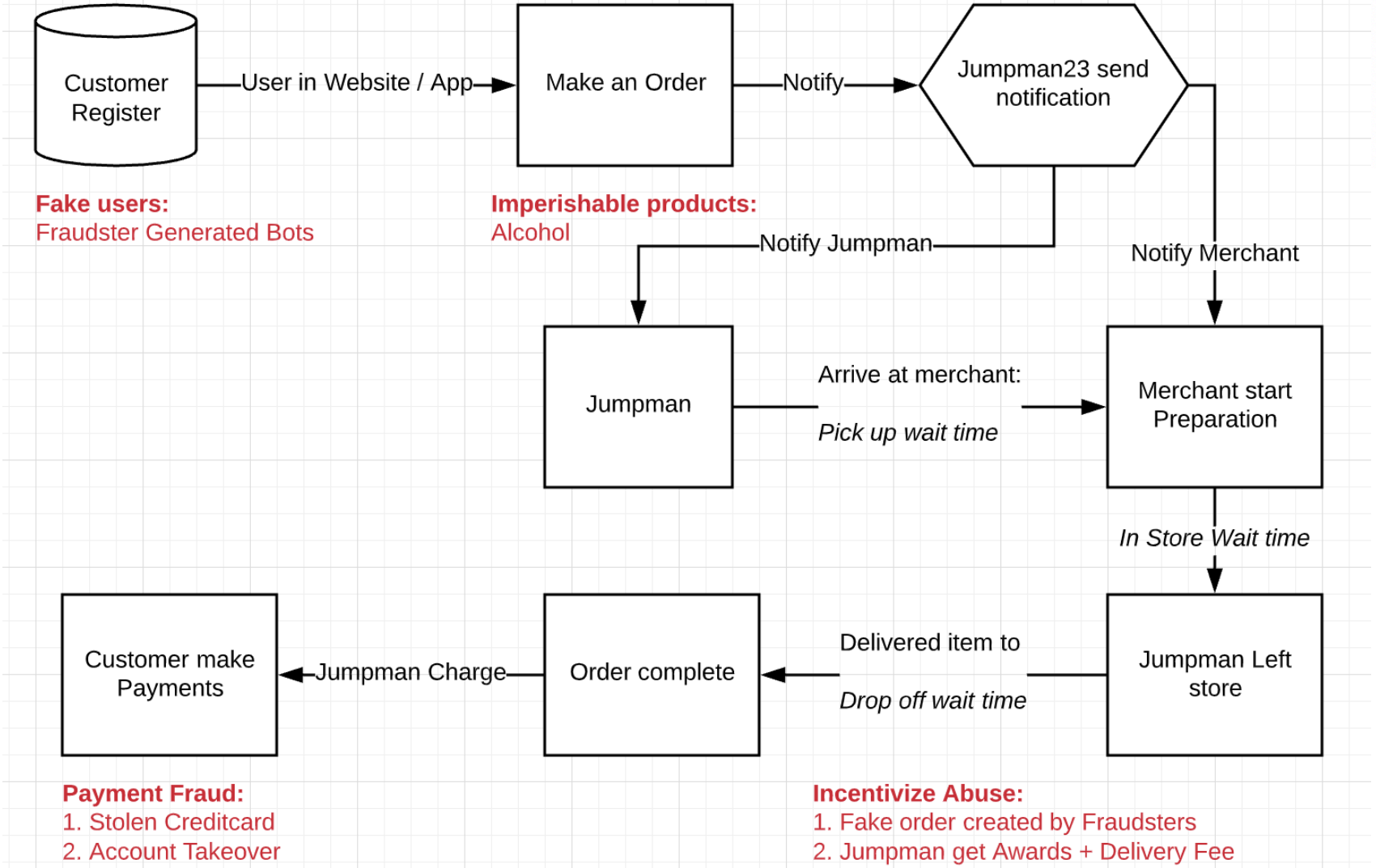
Issues

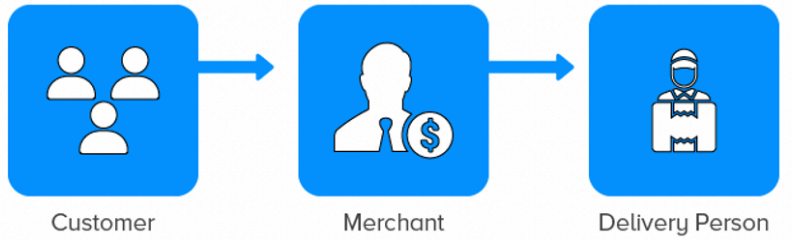
1. Duplicates
2. Missing values
3. Inappropriate value

Treatments

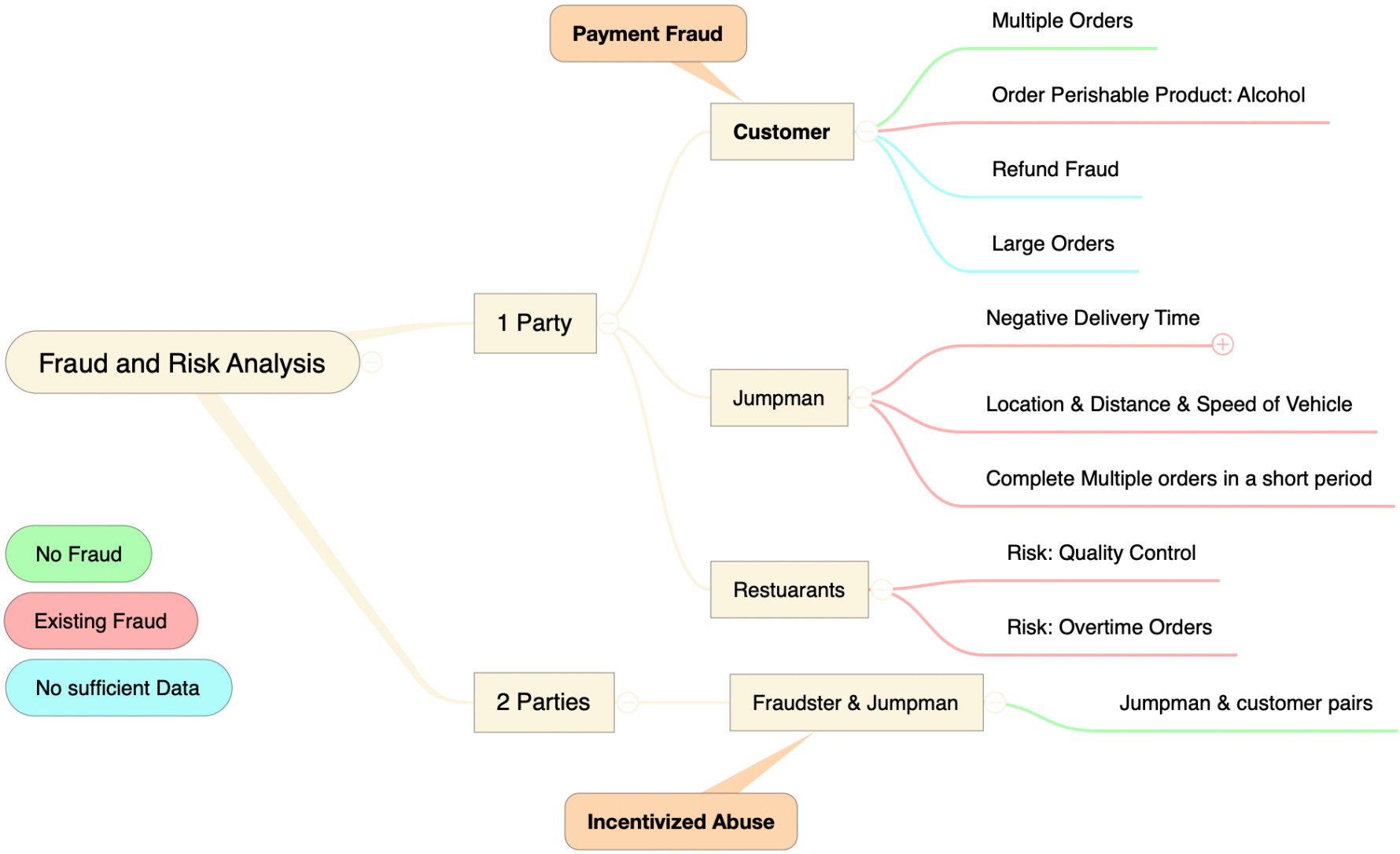
1. Unique Key (Database constraints)
2. Type of Missing values
   - Missing at Random vs
   - Missing not at Random
3. Inappropriate value (Value constraints)
   - Negative in store/arrival wait time
   - Investigate Jumpman
     - Proper training vs Fraud

Risk and Fraud analysis: the Food Delivery Ecosystem





Further Breakdown



Fraud and Risk Parties

Fraud Party	Hypothesis	How to Check it?
Jumpman	Within the time limit, completing the order is not realistic. e.g. Walker should not exceed 10 KM/hour	Feature for: Speed/ Distances Distance (Lat, Lon)
Customer	Customers take 3+ orders within an hour	Customers, order time
Customer	Customers who only order Perishable products	customer_id, item_category
Customer	The place is on the sea/ surface of the earth	Longitude, Latitude

Risk Party	Issues	How to improve?
Merchant	Average food preparation time is above 1 hour	Predictive an expected orders, prepare before order arrives.
Jumpman	Forget to check-in their arrival	Advice them / send notification

More robust Fraud-detection

Fraudulent Detection (History)

Objective: identify the likelihood or fraudulent activities among customers/ jumpman.

Cases I identified as Labels 0|1

Fraudulent Detection (Predictive)

Objective: predict the fraudulent activities before the fraud happens.

Data Need to Collect:

- Registration: GPS, user name
- App Download: Device ID, cookies
- Stolen Credentials Market

Combine with the Labels

Recommendation

1. **Fix the data integrity issues. Stored data in database with rules (constraints).**

2. **Use historical information to build a Fraud & Risk scorecard.**

3. **Use identified fraud patterns, build classification model to prevent future fraud activities.**

4. **Customer satisfaction**

   - **Advice merchant to prepare the items in advance.**
   - **Recommend customers with closest store. (Shake Shack case)**