Project Summary

There is an open project available on Kaggle, an online organization that connects analysts and businesses alike, where farmer's market data has been nationally collected from more than one source. There is a criticism that farmers markets are largely inaccessible to many Americans, especially of low socio-economic status. Let us say this is an ongoing case study and stakeholders will need to continually add information to the already existing flat csv files (2 are involved to be exact). There is a wide range of information available that describe different aspects of these farmers' markets. The USDA source includes pieces of information such as name and address, various online and social media platforms used for advertising, what kinds of produce can be found at each, forms of payment accepted, etc. The Wikipedia source includes details more so involving financial and demographic details of various geolocations surrounding the location of the markets.

The expected outcome would be to come up with a more sustainable means of storing and retrieving this information. Quite frankly its current state is quite messy. There are quite a few opportunities for important underlying aspects to be missed as there are a lot of free-form text columns that should not be inputted as such. There are also data inconsistency issues as some columns are storing values of similar context in different formats. As far as for data cleansing a lot of the heavy lifting has already been taken care of, so now the conversation must turn to solution longevity. The structure of these files needs to be broken down into separate categories, or entities, where attributes can be organically dissected and fitted into such. This would allow for faster turnaround time with analytical reports, and a more structured approach to future data entry instances.

Stakeholders

For the purpose of this project I would say the host of the project on Kaggle could be the stakeholder in this situation, as well as any other members involved in its creation. The administrator of this project will be the individuals involved the case study's continued expansion. They will need to make sure issues do not arise as new market information is entered into the infrastructure. Actual users of this database include the analysts investigating the organization's concerns around whether farmer's markets subtlety discriminate their audiences to be communities of higher socioeconomic status.

Business Rules

The stakeholders work to capture first general information about national farmer's markets:

- As mentioned, each market is designated their own unique ID, and are required to have their market's name available.
- Address, as well as latitude and longitude coordinates are also required.
- Season start and end dates and times are kept track of.
- There should no be duplicate entries for markets ID's.
- Each entry should be for a unique market.

Next Web URL's and social media URL's should be organized together as well:

- Website URL's can be shared by many regional markets. An instance of this could be thought of *Pure Michigan* which is an organizational website that connect people to various surrounding businesses and organizations.
- Markets have the option to enter social media URL's which includes FaceBook, Twitter, YouTube, Instagram, and other media platforms.
- Markets should only have one entry per form of media if they choose to list media URL's.

Farmers' markets payment information is captured:

- Farmer's markets also can provide what various payment forms they can accept.
- These are stored as yes or no answers.
- The options tracked are credit, WIC, WICcash, SFMNP, and SNAP. Each market can have multiple forms of payment available.

Market's also have many details about what kinds of products can be found at their markets:

- These are also stored as yes or no responses.
- Market's can have multiple kinds of products entered.
- They may also share whether their products are organic or wild harvested.
- Not every item has to be described as organic, wild harvested, or not; rather it is a general response to whether a market contains products of these kinds.
- There is also an option to enter the last time any of these pieces of information were last updated by the markets' themselves.

The stake holders additionally have captured information about various money and demographic metrics which are tracked by county and state:

- These metrics are required. Depending on the geological size of the market coverage the financial and demographic metrics available may be related to more than one market.
- One market may be related to several of these geological demographic and financial metrics.
- These also have their own unique ID's.
- Financial metrics include per capita, median household income, and median family income.
- Demographic information includes population and number of households.

Data Dictionary

Entity, Attribute, & Relationships Glossary:

Entity	Attribute
Market	FMID [r u] Name [r] Physical Address [r c] Coordinates [r c] SeasonStart [r c] SeasonEnd [r c] MarketInfoUpdate
Media	WebsiteURL FacebookURL TwitterURL InstagramURL
Payment	PaymentTypes [c]
Product	SourceTypes [c] ProductTypes [c]
Demographic	DemographicLocation [r c] PerCapita [r] MedianHouseholdIncome [r] MedianFamilyIncome [r] Population [r] NumberOfHouseholds [r]
Relationships	
*Markets have optional many website URLs's; website URL's have optional many markets Markets use optional 1 payment types options; payment types options are used by optional many markets Markets have optional 1 product types options; product types options are offered by optional many markets Markets have mandatory 1 demographics; demographics have optional many markets	

Data Questions

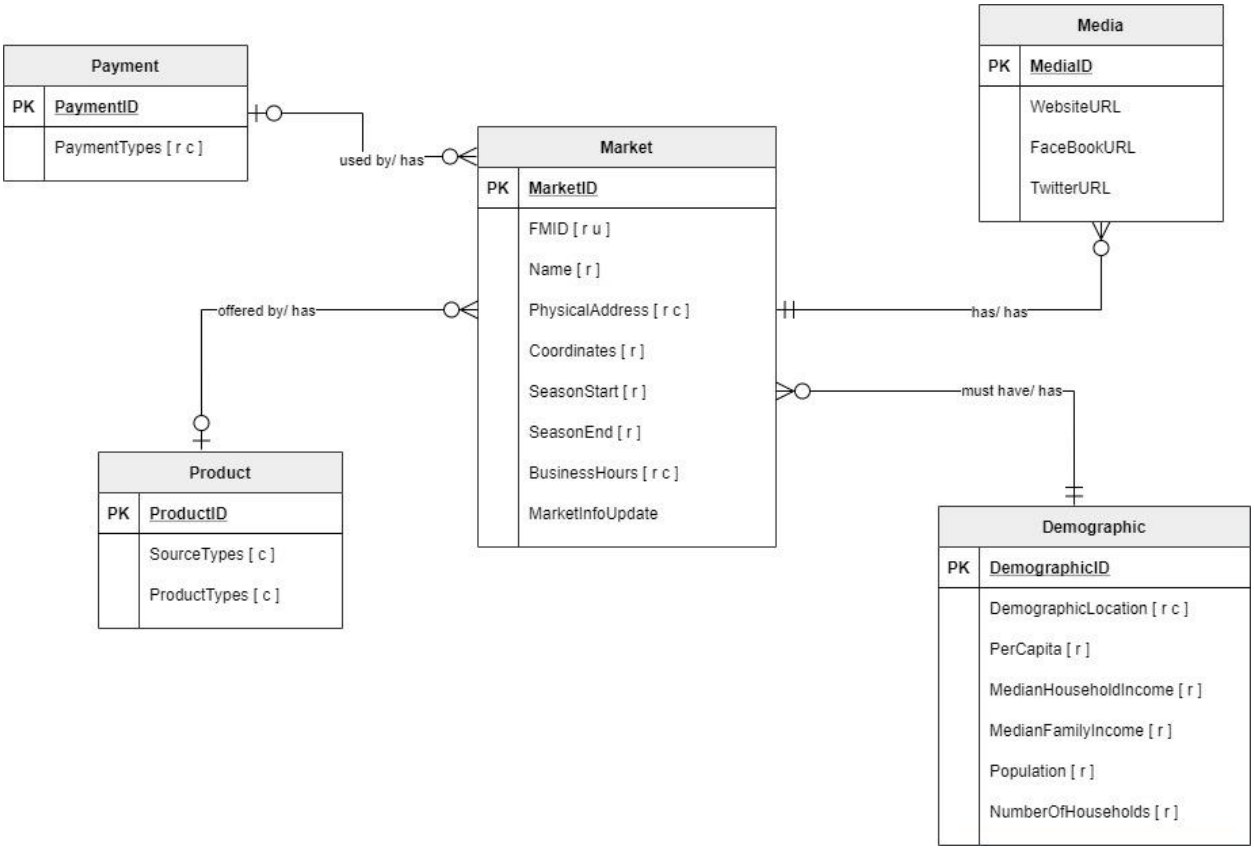
1. Based on the information provided are there indeed signs of socioeconomic segregation with respect to marketplace accessibility?
2. How do types of products available in a given market compare for an area of a higher socio-economic status against an area of a lower socio-economic status?
3. There are instances where multiple markets share the same address but occur on different days. Additionally, some markets may perceive to have different address, but share the same geographic coordinates. Are these indeed different markets? The answer to this question

could lead to an investigation as to why these different markets at the same location occur.

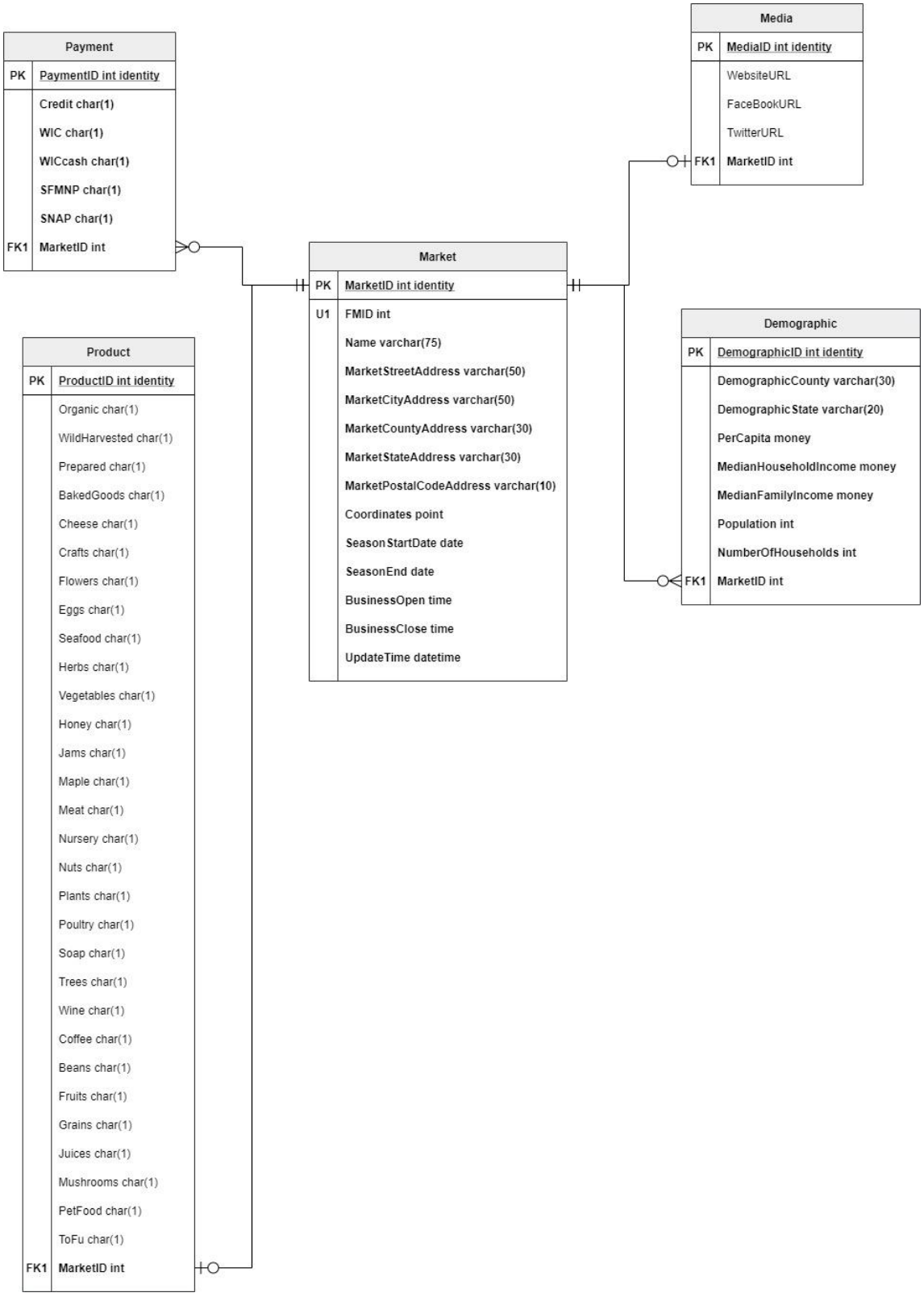
   a. Note: The data for this project endured some heavy cleaning, so instances of this inquiry may be few, but could lead to further investigation of the original files (perhaps even as a means of improving the overall database (i.e. relational model scope))

4. How do the market days, as well as open and close times differ when comparing areas of higher socio-economic status against areas of lower socio-economic status?
5. What's the relationship between accepted payment types for a given market and the market's associated demographics?
6. Are areas with higher socio-economic status more likely to have products sourced as "Organic", "WildHarvested", or "Prepared" by sellers than areas of lower socio-economic status?
7. There are many rows that do not offer websites or any form of media. What are the average economic metrics of the areas that do offer these forms of accessibility to communication with markets?

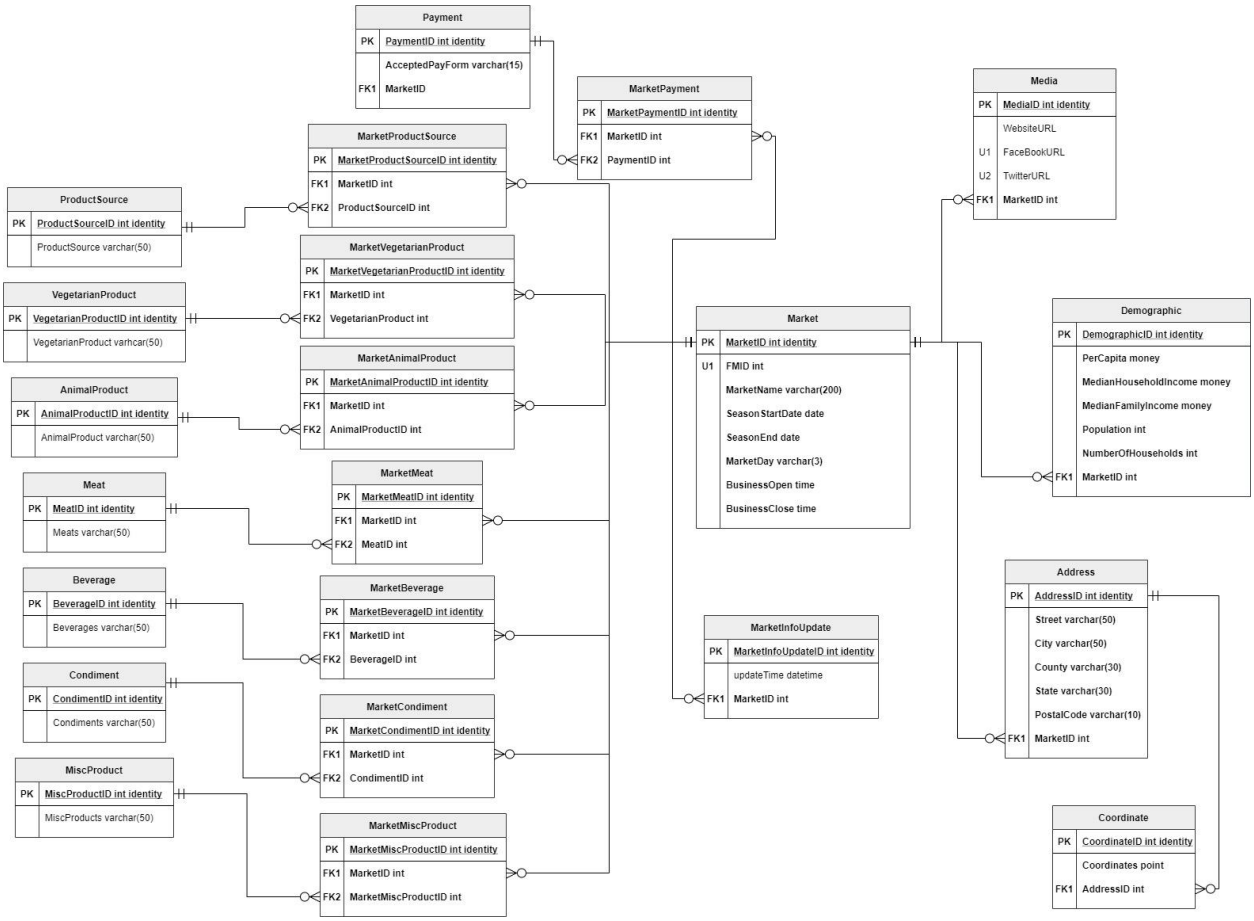
Conceptual Model



Logical Model[1]



Normalized Logical Model



[1] https://docs.microsoft.com/en-us/sql/relational-databases/spatial/point?view=sql-server-ver15

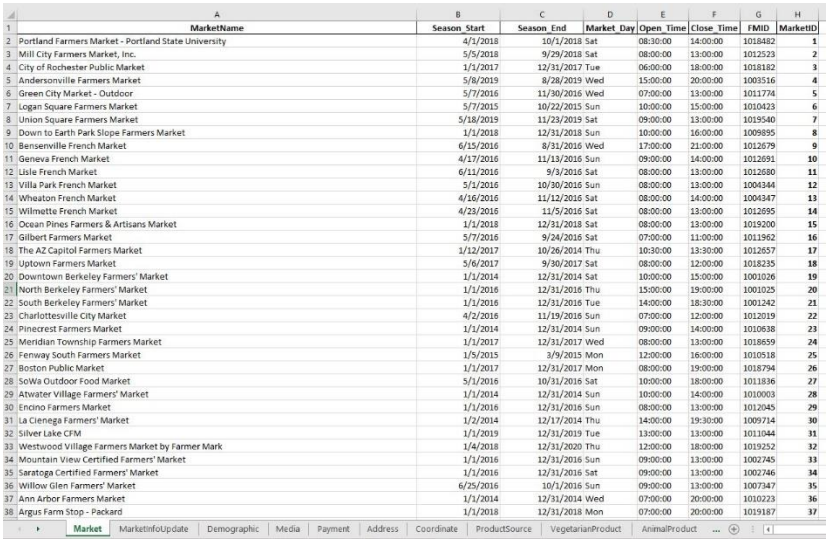
Developer Notes

Originally the USDA file portion of this table was very wide. Product types and sources were set up on a "yes" or "no" basis. From a scalability standpoint this would have cause design issues as new product types are introduced (essentially would require the creation of a new column in the RDBS for every instance of an introduction). It made more sense to melt these columns into row values (i.e. making the table tall). This decision changed the entire architecture of the original logical design as it introduced several new multivalued columns, which is why there is a stack of associative relations now. In doing so the final (joined) table ended up being over 7 million rows:

```
In [365]: FarmerWiki_Market_tall_clean.shape
Out[365]: (7164252, 33)
```

This caused relational development issues since Excel cannot read in that many rows, so all refinements made to the relations was performed in Python. After breaking apart the dataframe into the individual relations (including the associative relations) I was able to condense the tables based on their relationship with FMID or the relations' attribute (depending on the relation's relationship with "Market") since FMID values were controlled for as being unique and not null during the cleansing process. Associative relations were checked for duplicates in combination of FMID and the relation's respected attribute. This allowed for duplicate values to be removed while preserving "Market's" many-to-many relationship with its respect counterparts.

After completing these tasks the individual dataframe's were read into individual sheets of a newly created Excel spreadsheet. Creation of keys (primary & foreign) were completed in Excel. Foreign keys for the associative tables were imported from their associated tables using *VLOOKUP*.



	A	B	C	D	E	F	G	H
1	MarketName	Season_Start	Season_End	Market_Day	Open_Time	Close_Time	FMID	MarketID
2	Portland Farmers Market - Portland State University	4/1/2018	10/1/2018	Sat	08:30:00	14:00:00	1018482	1
3	Mill City Farmers Market, Inc.	5/5/2018	9/29/2018	Sat	08:00:00	13:00:00	1012523	2
4	City of Rochester Public Market	1/1/2017	12/31/2017	Tue	06:00:00	18:00:00	1018182	3
5	Andersonville Farmers Market	5/8/2019	8/28/2019	Wed	15:00:00	20:00:00	1003516	4
6	Green City Market - Outdoor	5/7/2016	11/30/2016	Wed	07:00:00	13:00:00	1011774	5
7	Logan Square Farmers Market	5/7/2015	10/22/2015	Sun	10:00:00	15:00:00	1010423	6
8	Union Square Farmers Market	5/18/2019	11/23/2019	Sat	09:00:00	13:00:00	1019540	7
9	Down to Earth Park Slope Farmers Market	1/1/2018	12/31/2018	Sun	10:00:00	16:00:00	1009895	8
10	Bensenville French Market	6/15/2016	8/31/2016	Wed	17:00:00	21:00:00	1012679	9
11	Geneva French Market	4/17/2016	11/13/2016	Sun	09:00:00	14:00:00	1012691	10
12	Lisle French Market	6/11/2016	9/3/2016	Sat	08:00:00	13:00:00	1012680	11
13	Villa Park French Market	5/1/2016	10/30/2016	Sun	08:00:00	13:00:00	1004344	12
14	Wheaton French Market	4/16/2016	11/12/2016	Sat	08:00:00	14:00:00	1004347	13
15	Wilmette French Market	4/23/2016	11/5/2016	Sat	08:00:00	13:00:00	1012695	14
16	Ocean Pines Farmers & Artisans Market	1/1/2018	12/31/2018	Sat	08:00:00	13:00:00	1019200	15
17	Gilbert Farmers Market	5/7/2016	9/24/2016	Sat	07:00:00	11:00:00	1011962	16
18	The AZ Capitol Farmers Market	1/12/2017	10/26/2014	Thu	10:30:00	13:30:00	1012657	17
19	Uptown Farmers Market	5/6/2017	9/30/2017	Sat	08:00:00	12:00:00	1018235	18
20	Downtown Berkeley Farmers' Market	1/1/2014	12/31/2014	Sat	10:00:00	15:00:00	1001026	19
21	North Berkeley Farmers' Market	1/1/2016	12/31/2016	Thu	15:00:00	19:00:00	1001025	20
22	South Berkeley Farmers' Market	1/1/2016	12/31/2016	Tue	14:00:00	18:30:00	1001242	21
23	Charlottesville City Market	4/2/2016	11/19/2016	Sun	07:00:00	12:00:00	1012019	22
24	Pinecrest Farmers Market	1/1/2014	12/31/2014	Sun	09:00:00	14:00:00	1010638	23
25	Meridian Township Farmers Market	1/1/2017	12/31/2017	Wed	08:00:00	13:00:00	1018659	24
26	Fenway South Farmers Market	1/5/2015	3/9/2015	Mon	12:00:00	16:00:00	1010518	25
27	Boston Public Market	1/1/2017	12/31/2017	Mon	08:00:00	19:00:00	1018794	26
28	SoWa Outdoor Food Market	5/1/2016	10/31/2016	Sat	10:00:00	18:00:00	1011836	27
29	Atwater Village Farmers' Market	1/1/2014	12/31/2014	Sun	10:00:00	14:00:00	1010003	28
30	Encino Farmers Market	1/1/2016	12/31/2016	Sun	08:00:00	13:00:00	1012045	29
31	La Cienega Farmers' Market	1/2/2014	12/17/2014	Thu	14:00:00	19:30:00	1009714	30
32	Silver Lake CFM	1/1/2019	12/31/2019	Tue	13:00:00	13:00:00	1011044	31
33	Westwood Village Farmers Market by Farmer Mark	1/4/2018	12/31/2020	Thu	12:00:00	18:00:00	1019252	32
34	Mountain View Certified Farmers' Market	1/1/2016	12/31/2016	Sun	09:00:00	13:00:00	1002745	33
35	Saratoga Certified Farmers' Market	1/1/2016	12/31/2016	Sat	09:00:00	13:00:00	1002746	34
36	Willow Glen Farmers' Market	6/25/2016	10/1/2016	Sun	09:00:00	13:00:00	1007347	35
37	Ann Arbor Farmers Market	1/1/2014	12/31/2014	Wed	07:00:00	20:00:00	1010223	36
38	Argus Farm Stop - Packard	1/1/2018	12/31/2018	Mon	07:00:00	20:00:00	1019187	37

Market | MarketInfoUpdate | Demographic | Media | Payment | Address | Coordinate | ProductSource | VegetarianProduct | AnimalProduct | ...