

Project Name, Participants, and Workflow:

Project Name:

Time Series Analysis of Iowa Liquor Sales

Participants:

Andre Sharp 11374558 AndreSharp@my.unt.edu

Dan Waters 11457837 DanWaters@my.unt.edu

Bryan Adams 10976617 BryanAdams@my.unt.edu

Haiyi Wang 11528159 HaiyiWang@my.unt.edu

Workflow:

We will use Discord, email, and Zoom meetings to connect with each other. We will also have the weekly meeting on Saturday before the class.

We will use GitHub as the code version control. We can update the code and version on GitHub.

Project Abstract:

Motivation:

As we all know, retail businesses have a ‘goldilocks’ problem when it comes to inventory: don’t stock too much, but don’t stock too little. If we stock too much, it will occupy the huge amount of money and have the risk that products can’t be sold due to the production time, consumer favor, and so on. If we stock too little, we will probably miss the time to sell the product and miss the big chance to earn money. So, we need to use historic data and suitable models to analyze and to predict the consumption of products in the future. It will help the business owner to make a good decision in advance.

Objectives:

Our project is to use Iowa liquor sales data in multiple ways to analyze customer behavior for liquor consumption. As a brief bit of background, Iowa is an alcohol beverage control state, meaning that the state maintains a monopoly on wholesaling of alcohol throughout the State. Effectively, private retailers must purchase their alcohol from the state before selling it to individual consumers. We will use a time series model (ARIMA) to train the historic data and forecast the consumption in the future. All the computation and model deployment are on the google cloud. We will use Google BigQuery ML to extract the data and create a dashboard to show predicted metrics.

Google Cloud

csoc5214 p1

Explorer

+ ADD DATA

sales

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

Type to search

Viewing pinned projects.

csoc5214 p1

External connections

csoc5214.love_sales

sales

SCHEMA

DETAILS

PREVIEW

Row	invoice_and_item_number	date	store_number	store_name	address	city	zip_code	store_location
1	506218300004	2012-10-09	1249	Hy-Vee Food Store / Indiana	910 N JEFFERSON	INDIANOLA	50125	null
2	524912700069	2015-04-06	3335	Hy-Vee Food Store #1 / WDM	1700 VALLEY WEST DR	WEST DES MOINES	50265	
3	INV-4542320005	2022-03-09	2675	Hy-Vee #2 / Coralville	3283 Crosspark Rd	Coralville	52241-0	POINT (-91.605881, 41.722419)
4	INV-4567990092	2022-03-17	5102	Wilkie Liquors	724 1st Street NE	Mount Vernon	52314-0	POINT (-91.410401, 41.918328)
5	INV-14785100049	2018-10-02	2619	Hy-Vee Wine and Spirits / WDM	1725 74th St	West Des Moines	50266-0	POINT (-93.808855, 41.598155)
6	INV-2966030023	2020-08-21	5857	EZ Stop 1 - Dubuque	700 Rhomburg Avenue	Dubuque	52001-0	POINT (-90.663442, 42.51732)
7	531619300022	2016-04-05	4236	Fareway Stores #551 / Eagle G	205 NW 1ST ST	EAGLE GROVE	50533	POINT (-93.903801, 42.665165)
8	INV-34828300007	2021-03-08	4969	Lake Liquors Wine and Spirits	910 N 8th St W	Clear Lake	50428-0	POINT (-93.396651, 43.142775)
9	INV-35048900015	2021-03-16	3814	Costco Wholesale #788 / WDM	7205 Mills Civic Pkwy	West Des Moines	50266-0	POINT (-93.806489, 41.561342)
10	514422100012	2013-09-09	4495	Cassey's General Store #3055 /	504 G AVE	GRUNDY CENTER	50638	POINT (-92.770411, 42.361494)
11	525892800112	2015-05-28	5102	Wilkie Liquors	724 1st ST SE	MOUNT VERNON	52314	POINT (-91.410401, 41.918328)
12	527828100067	2015-09-10	5162	Urbankade Liquor	6401 DOUGLAS AVE	URBANDALE	50322	POINT (-93.705732, 41.629411)
13	504061800048	2012-02-14	4167	lowe's Market Inc.	1256 IOWA ST	DUBUQUE	50201	POINT (-90.668138, 42.504959)
14	525708200151	2015-05-18	2633	Hy-Vee #3 / BDI / Des Moines	3020 321 S 14TH ST	DES MOINES	50320	POINT (-93.596754, 41.554010)
15	515488000076	2013-11-04	3385	Sams Club 8162 / Cedar Rapids	2605 BLAIRS FERRY RD NE	CEDAR RAPIDS	52402	POINT (-91.67969, 42.031819)
16	509202700061	2012-11-28	2200	Sac Liquor Store	619 E MAIN ST	SAC CITY	50583	POINT (-94.974011, 42.421341)
17	INV-21967900004	2019-09-17	5715	Raymarket	3432 Lafayette Road	Evansdale	50707	POINT (-92.291731, 42.478437)
18	INV-15132900015	2018-10-18	3420	Sams Club 6344 / Windsor Hl	1101 73rd St	Windsor Heights	50311-0	POINT (-93.718027, 41.599172)
19	INV-31622700005	2020-11-03	2338	Hy-Vee Food Store #3 / Waterloo	1422 Flammang Dr	Waterloo	50702-0	POINT (-92.327917, 42.459938)
20	508503000001	2012-10-23	4204	Fareway Stores #025 / Clinton	1350 11TH ST NW	CLINTON	52733	POINT (-90.211315, 41.860978)
21	528146200009	2015-09-28	4166	Jesup Food Center	1314 7TH ST	JESUP	50648	POINT (-92.062259, 42.469327)
22	510742700085	2013-02-21	2513	Hy-Vee Food Store #2 / Iowa Cl	812 S 1ST AVE	IOWA CITY	52240	null
23	525603100027	2015-05-12	2666	Hy-Vee #2 / Ankeny	2510 SW STATE ST	ANKENY	50023	POINT (-93.621824, 41.705188)
24	INV-07998000015	2017-10-03	3625	Wal-Mart 0892 / Ankeny	1002 SE National Dr	Ankeny	50021	POINT (-93.582348, 41.704864)
25	INV-15288700115	2018-10-25	4829	Central City 2	1501 Michigan Ave	Des Moines	50314-0	POINT (-93.613739, 41.60572)
26	615648000001	2015-11-13	7040	Hy-Vee Wine and Spirits / I am	1301 19TH AVE NW	EMERY	51101	POINT (-94.18355000000000, 4

Results per page: 50 1 - 50 of 24229431

Programming language:

Cloud platform:

Dataset:

Features:

We have 24 features in our dataset, The datatype of feature are date, string, integer and float.

The screenshot shows the Google Cloud BigQuery interface. At the top, there's a search bar and navigation icons. Below that, the 'sales' table is selected, and the 'SCHEMA' tab is active. The schema table lists various fields with their data types and descriptions.

Field Name	Data Type	Nullable	Description
date	DATE	NULLABLE	Date of order
store_number	STRING	NULLABLE	Unique number assigned to the store who ordered the liquor.
store_name	STRING	NULLABLE	Name of store who ordered the liquor.
address	STRING	NULLABLE	Address of store who ordered the liquor.
city	STRING	NULLABLE	City where the store who ordered the liquor is located
zip_code	STRING	NULLABLE	Zip code where the store who ordered the liquor is located
store_location	STRING	NULLABLE	Location of store who ordered the liquor. The Address, City, State and Zip Code are geocoded to provide geographic coordinates. Accuracy of geocoding is dependent on how well the address is interpreted and the completeness of the reference data used.
county_number	STRING	NULLABLE	Iowa county number for the county where store who ordered the liquor is located
county	STRING	NULLABLE	County where the store who ordered the liquor is located
category	STRING	NULLABLE	Category code associated with the liquor ordered
category_name	STRING	NULLABLE	Category of the liquor ordered.
vendor_number	STRING	NULLABLE	The vendor number of the company for the brand of liquor ordered
vendor_name	STRING	NULLABLE	The vendor name of the company for the brand of liquor ordered
item_number	STRING	NULLABLE	Item number for the individual liquor product ordered.
item_description	STRING	NULLABLE	Description of the individual liquor product ordered.
pack	INTEGER	NULLABLE	The number of bottles in a case for the liquor ordered
bottle_volume_ml	INTEGER	NULLABLE	Volume of each liquor bottle ordered in milliliters.
state_bottle_cost	FLOAT	NULLABLE	The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered
state_bottle_retail	FLOAT	NULLABLE	The amount the store paid for each bottle of liquor ordered
bottles_sold	INTEGER	NULLABLE	The number of bottles of liquor ordered by the store
sale_dollars	FLOAT	NULLABLE	Total cost of liquor order (number of bottles multiplied by the state bottle retail)
volume_sold_liters	FLOAT	NULLABLE	Total volume of liquor ordered in liters. (i.e. (Bottle Volume (ml) x Bottles Sold)/1,000)
volume_sold_gallons	FLOAT	NULLABLE	Total volume of liquor ordered in gallons. (i.e. (Bottle Volume (ml) x Bottles Sold)/3785.411784)

Date: date of the order (eg: '2020-08-19')

Store_name: name of the store (eg: Wilkie Liquors)

City: the city of the store (eg: Mount Vernon)

Category_name : the category of the liquor (eg: Neutral Grain Spirits)

Vendor_name : the name of the vendor (eg: LUXCO INC)

Item_description: description of the individual liquor (eg: Templeton Rye)

State_bottle_cost: the cost of each bottle of liquor

State_bottle_retail: the retail price of each bottle of liquor

Bottles_sold: the amount of liquor sold

Sale_dollars: the amount of money of liquor sales.

Models:

We will use the time-series Model (ARIMA) in our project. In our daily life, Autoregressive Integrated Moving Average (ARIMA) models have many places to use. For example, if we need forecast the stock of the product, we can use this model to forecast. In some factory, the stock of product is very important for them to plan to product. If we could use ARIMA model to predict the stock in the future, it would be helpful for us.

ARIMA models are a general class of models used for forecasting time series data. ARIMA models are generally denoted as ARIMA (p,d,q) where p is the order of autoregressive model, d is the degree of differencing, and q is the order of moving-average model. ARIMA models use differencing to convert a non-stationary time series into a stationary one, and then predict future values from historical data.

Milestones:

Week 1 Sept 3:

1. Project selection

Week 2 Sept 10:

1. Group formed
2. Make the work plan and assignment
3. Complete the project proposal

Week 3 Sept 17:

1. Create google cloud account and set up
2. Extract the data from dataset and transfer to Google Cloud
3. Preprocessing and cleaning the data
4. Visualization the data and EDA

Week 4 Sept 24:

1. Deploy the time-series model and training the data
2. Forecasting the data
3. Create the dashboard

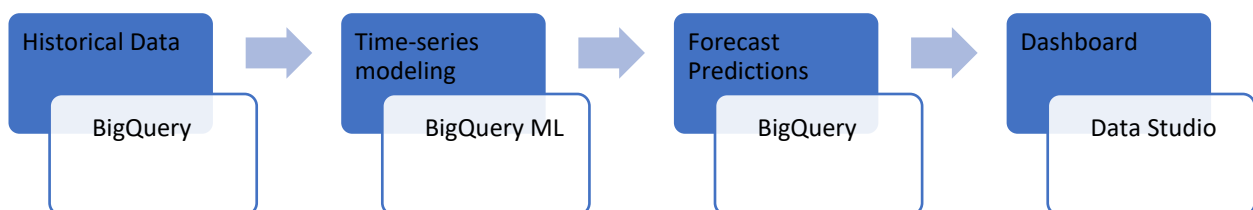
Week 5 Oct 1:

1. Create the pipeline
2. Testing the system

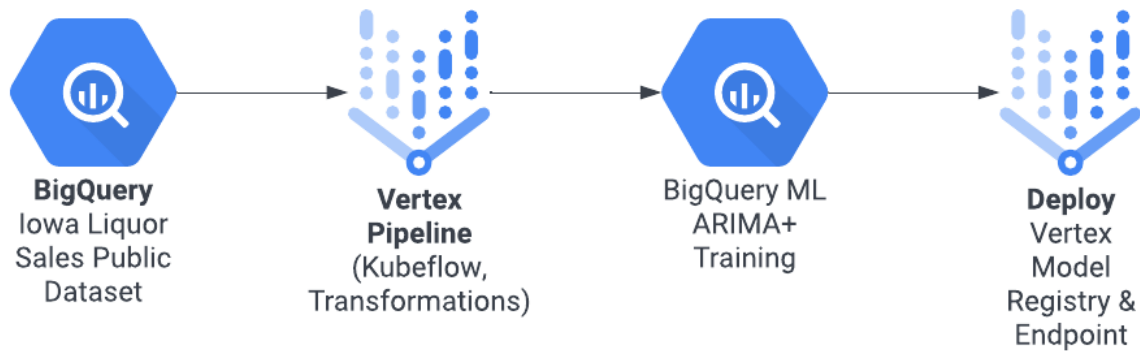
Week 6 Oct 8:

1. Complete the final report
2. Prepare the presentation

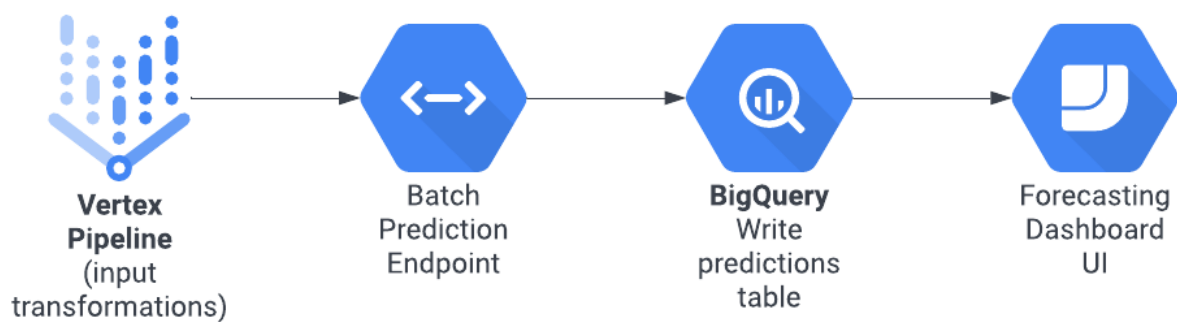
Architecture of the system



Training pipeline



Serving pipeline (batch)



Resources and Related Projects:

[1] Iowa liquor sales data

<https://data.iowa.gov/Sales-Distribution/Iowa-Liquor-Sales/m3tr-ghgy>

[2] ARIMA

<https://www.capitalone.com/tech/machine-learning/understanding-arima-models/>

[3] BigQuery ML on Google Cloud

<https://cloud.google.com/blog/topics/developers-practitioners/how-build-demand-forecasting-models-bigquery-ml>