Let’s take a demo customers’ expenses data. We will work on data cleaning, aggregation.

Once the data is in the form we want, we will visualize it on web with the help of NextJS.

Data Import

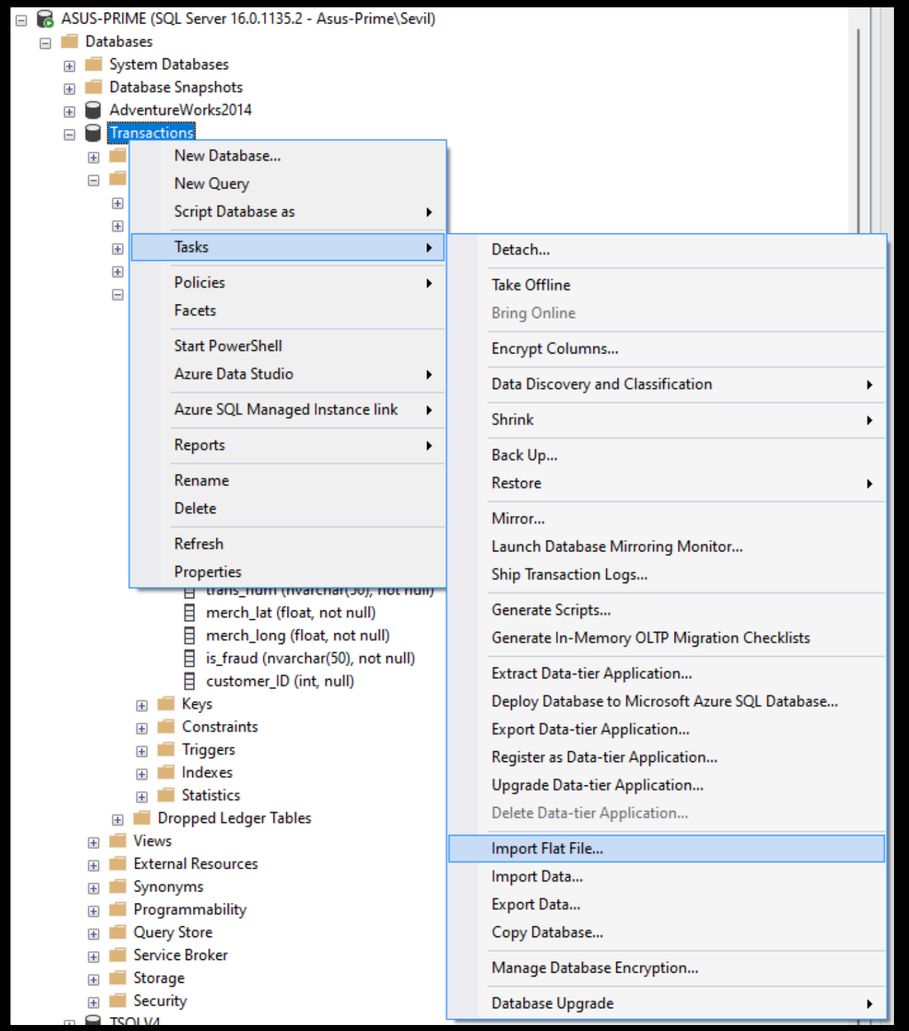
- Create a database named **Transactions**

```
CREATE DATABASE Transactions;
```

- To import CSV File into **Transactions** table

```
Right click on Transaction database ->
Tasks -> Import Flat File

In opened screen specify location of file ->
Preview your data -> Modify columns ->
Next and Finish
```



	trans_date_trans_time	merchant	category	amt	city	state	lat	long	city_pop	job	dob	trans_num
13	5/1/2019 22:18	"Hauck, Dietrich and Funk"	kids_pets	4.58	Wales	AK	64.7556	-165.6723	145	"Administrator, education"	9/11/1939	8fa7880cf01e6adc96e9654d69f3f035
14	5/1/2019 22:32	Pouros-Haag	shopping_pos	730.78	Wales	AK	64.7556	-165.6723	145	"Administrator, education"	9/11/1939	2396a5b8e277a4ce1d4fa939d6e7b19b
15	5/1/2019 22:33	Goyette Inc	shopping_net	1006.4	Wales	AK	64.7556	-165.6723	145	"Administrator, education"	9/11/1939	4d7e567247b6c4529ce4c32c03b2f040
16	5/1/2019 22:38	"Baumbach, Strosin and Nicolas"	shopping_pos	830.72	Wales	AK	64.7556	-165.6723	145	"Administrator, education"	9/11/1939	773a3305db095657e1860a39dda24cf2
17	5/1/2019 23:17	Pacocha-O'Reilly	grocery_pos	311.92	Wales	AK	64.7556	-165.6723	145	"Administrator, education"	9/11/1939	191b3dcec7a6a49438be93cfbeaea9c1
18	5/1/2019 23:26	Barrows PLC	shopping_pos	762.93	Browning	MO	40.029	-93.1607	602	Cytogeneticist	14-07-1954	19b126ecf4c79997e9edc6f260c09880
19	6/1/2019 18:39	Fisher-Schowalte	shopping_net	855.88	Browning	MO	40.029	-93.1607	602	Cytogeneticist	14-07-1954	bbae703c3794b7738d08739afbaef898
20	6/1/2019 23:33	Gleason-Macejkovic	shopping_net	909.29	Browning	MO	40.029	-93.1607	602	Cytogeneticist	14-07-1954	b98eb7183ee8a48e83539addc80ea9cc
21	12/1/2019 1:46	Kutch LLC	gas_transport	7.04	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	436937887b481bbb237956bb4e592193
22	12/1/2019 15:36	Kihn Inc	shopping_pos	644.4	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	c50735baa9016d1707cd945c5926f687
23	12/1/2019 23:09	Kris-Weimann	misc_net	784.77	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	5900fb4411f88bda3f613fb8fa20f925
24	12/1/2019 23:57	"Kihn, Abernathy and Douglas"	shopping_net	1116.8	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	cad6a618cba464438b5566597933f870
25	13-01-2019 00:14	Kiehn-Emmerich	grocery_pos	320.6	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	98649f992d93377c62c09e51a1bd2cc9
26	13-01-2019 01:59	"Osinski, Ledner and Leuschke"	grocery_pos	276.8	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	a28b688a63d93fb5935d6a3dfb22ecf6
27	13-01-2019 02:16	Vandervort-Funk	grocery_pos	336.2	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	62e0c792222afcc2ef1644a81ffdeb47
28	13-01-2019 21:10	Terry-Huel	shopping_net	994.92	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	a2c0f59146a22479dd65bae74d6d7741
29	13-01-2019 22:01	Volkman-Predovic	shopping_net	1062.95	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	2182d9659d60dfa68fa1f39baa2fc635
30	13-01-2019 22:17	"Roob, Conn and Tremblay"	shopping_pos	770.94	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	0e4cd5fbfa6845c164b8bb7a6b21ab14
31	13-01-2019 23:04	"Tillman, Fritsch and Schmitt"	misc_net	816.89	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	186c2aeb244b826107f83b307233fb08
32	13-01-2019 23:17	Larson-Moen	entertainment	510.59	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	f1edc60904bafa8aac00a0f5e9026d0c
33	13-01-2019 23:28	Kutch and Sons	grocery_pos	307.5	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	9bc5cb494abc3af2b02ca33e0d076f74
34	13-01-2019 23:45	"Parisian, Schiller and Altenwer...	misc_net	790.35	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	7d7d61dc3b301c78ca3c0cf73e8ed72e
35	13-01-2019 23:57	Kozey-Boehm	shopping_net	1065.43	Ravenna	NE	41.0233	-98.9041	2202	"Solicitor, Scotland"	21-06-1974	049087fe5d27b77c7238fa46bb18c99d
36	14-01-2019 00:27	Rau and Sons	grocery_pos	306.81	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	fdc202f9f1dd556a51775c6d8060c58d
37	14-01-2019 00:37	Parisian and Sons	gas_transport	12.57	San Jose	CA	37.3304	-121.7913	973849	Science writer	12/6/1955	dec7f564c518a3f5878016461d766ffa

Data Cleaning
PART 1

In **trans_date_trans_time** and **dob** columns some dates given with " - ", others given with " / " . We will change all of them with " - ".

```
UPDATE Purchases
SET trans_date_trans_time = REPLACE(trans_date_trans_time, '/', '-')
WHERE trans_date_trans_time LIKE '%/%';

UPDATE Purchases
SET dob = REPLACE(dob, '/', '-')
WHERE dob LIKE '%/%';

SELECT trans_date_trans_time FROM Purchases;
```

	trans_date_trans_time
37	14-01-2019 00:37
38	14-01-2019 00:46
39	14-01-2019 01:41
40	14-01-2019 01:42
41	14-01-2019 02:27
42	14-01-2019 08:44
43	14-01-2019 09:30
44	14-01-2019 16:15
45	14-01-2019 22:28
46	14-01-2019 23:17
47	14-01-2019 23:18
48	15-01-2019 01:07
49	15-01-2019 22:50
50	15-01-2019 23:01
51	15-01-2019 23:12
52	17-01-2019 01:30
53	17-01-2019 08:59
54	17-01-2019 22:26
55	17-01-2019 22:45
56	18-01-2019 00:23
57	18-01-2019 03:36
58	18-01-2019 03:54
59	18-01-2019 14:38
60	18-01-2019 22:06
61	18-01-2019 22:12

Data Cleaning
PART 2

In **merchant** and **job** columns some strings used “ , ” and “ ‘ ” . We will erase these characters.

```
UPDATE Purchases
SET merchant = REPLACE(merchant, '"', '')
WHERE merchant LIKE '%"%';

UPDATE Purchases
SET job = REPLACE(job, ',', '')
WHERE job LIKE '%,%';

SELECT merchant, job FROM Purchases;
```

	merchant	job
1	Stokes Christiansen and Sipes	Administrator education
2	Predovic Inc	Administrator education
3	Wisozk and Sons	Administrator education
4	Murray-Smitham	Administrator education
5	Friesen Lt	Administrator education
6	Raynor Reinger and Hagenes	Administrator education
7	Heller-Langosh	Administrator education
8	Padberg-Welch	Cytogeneticist
9	McGlynn-Heathcote	Administrator education
10	Dooley-Thompson	Administrator education
11	Gottlieb Considine and Schultz	Cytogeneticist
12	Moen Reinger and Murphy	Administrator education
13	Hauck Dietrich and Funk	Administrator education
14	Pouros-Haag	Administrator education
15	Goyette Inc	Administrator education
16	Baumbach Strosin and Nicolas	Administrator education
17	Pacocha-O'Reilly	Administrator education
18	Barrows PLC	Cytogeneticist
19	Fisher-Schowalte	Cytogeneticist
20	Gleason-Macejkovic	Cytogeneticist
21	Kutch LLC	Solicitor Scotland
22	Kihn Inc	Solicitor Scotland
23	Kris-Weimann	Solicitor Scotland
24	Kihn Abernathy and Douglas	Solicitor Scotland
25	Kiehn-Emmerich	Solicitor Scotland

Modify column

- Let’s modify **trans_date_trans_time** column name to **trans_datetime**:

```
EXEC sp_rename 'Purchases.trans_date_trans_time',  'trans_datetime', 'COLUMN';
```

- Modify **trans_datetime** column data type from **String** to **DateTime**:

```
ALTER TABLE Purchases
ALTER COLUMN trans_datetime DATETIME;

UPDATE Purchases
SET trans_datetime = CONVERT(DATETIME, trans_datetime, 103);
```

Modify column

- Also modify **dob** column data type from **String** to **DateTime** :

```
ALTER TABLE Purchases
ALTER COLUMN dob DATE;

UPDATE Purchases
SET dob = CONVERT(DATE, dob, 103);
```

Add column

We need unique **customer_id** column for filtering purchases by user.

I divide this process to 2 steps:

- How can I detect different users ?
- How to define them unique ids ?

Step 1

Analyze data help of this command:

```
SELECT [dob] FROM Purchases
GROUP BY dob;
```

	dob
177	1946-05-28
178	1976-02-26
179	1987-10-28
180	1936-12-23
181	1986-02-17
182	1938-08-07
183	1973-05-07
184	1979-01-26
185	1974-06-21
186	1985-08-21
187	1967-01-24

```
SELECT [dob], [job] FROM Purchases
GROUP BY dob, job;
```

	dob	job
177	1973-05-16	Administrator local government
178	1987-05-19	Immigration officer
179	1972-04-18	Land/geomatics surveyor
180	1997-01-18	Geoscientist
181	1981-10-24	Contractor
182	1951-01-15	Therapist occupational
183	1993-11-24	Economist
184	1941-09-30	Accountant chartered
185	1957-07-27	Research scientist (medical)
186	1951-11-08	Dealer
187	1985-03-31	Land/geomatics surveyor

Step 2

Create and assign random ID to customer_id column by unique dob

```
ALTER TABLE Purchases ADD customer_ID int;

WITH Customer AS (
    SELECT DISTINCT
           dob,
           ABS(CHECKSUM(NEWID())) % 1000000 AS customer_id
    FROM Purchases
)

UPDATE Purchases
SET Purchases.customer_id = Customer.customer_id
FROM Purchases
INNER JOIN Customer ON Purchases.dob = Customer.dob;
```

	trans_datetime	merchant	category	amt	city	state	lat	long	city_pop	job	dob	trans_num	customer_ID
1	2019-01-01 00:00:00...	Lind-Buckridge	entertainment	220.11	Malad City	ID	42.1808	-112.262	4154	Nature conservation officer	1962-01-19	a1a22d7...	142414
2	2019-01-01 00:07:00...	Kiehn Inc	grocery_pos	96.29	Grenada	CA	41.6125	-122.5258	589	Systems analyst	1945-12-21	413636e...	761701
3	2019-01-01 00:09:00...	Beier-Hyatt	shopping_pos	7.77	High Rolls Mountain Park	NM	32.9396	-105.8189	899	Naval architect	1967-08-30	8a6293af...	241751
4	2019-01-01 00:21:00...	Bruen-Yost	misc_pos	6.85	Freedom	WY	43.0172	-111.0292	471	Education officer museum	1967-08-02	f3c43d33...	176084
5	2019-01-01 00:22:00...	Kunze Inc	grocery_pos	90.22	Honokaa	HI	20.0827	-155.488	4878	Physiotherapist	1966-12-03	95826e3...	266669
6	2019-01-01 00:22:00...	Nitzsche Kessler and Wol	shopping_pos	4.02	Valentine	NE	42.8062	-100.6215	4005	Network engineer	1945-03-15	20490f3f...	272690
7	2019-01-01 00:22:00...	Kihn Abernathy and Douglas	shopping_net	3.66	Westfir	OR	43.7575	-122.481	597	Forensic psychologist	1961-05-19	870c92b...	296368
8	2019-01-01 00:31:00...	Ledner-Pfannerstill	gas_transport	102.13	Thompson	UT	38.9999	-109.615	46	Surveyor minerals	1987-04-23	47238da...	160716
9	2019-01-01 00:34:00...	Stracke-Lemke	grocery_pos	83.07	Conway	WA	48.34	-122.3456	85	Research officer political party	1984-09-01	9b7a061...	825180
10	2019-01-01 00:40:00...	Cummerata-Jones	gas_transport	70.53	Athena	OR	45.8289	-118.4971	1302	Dealer	1976-10-18	7b91bd2...	52739
11	2019-01-01 00:41:00...	Huel-Langworth	misc_net	177.57	Thompson	UT	38.9999	-109.615	46	Surveyor minerals	1987-04-23	048dc7fc...	160716
12	2019-01-01 00:46:00...	Ferry Lynch and Kautze	misc_net	2.76	San Jose	CA	37.3304	-121.7913	973849	Science writer	1955-06-12	70ca7fe4...	672949
13	2019-01-01 00:49:00...	Little Gutmann and Lynch	shopping_net	83.52	Ravenna	NE	41.0233	-98.9041	2202	Solicitor Scotland	1974-06-21	5f7aafc8...	655422
14	2019-01-01 00:56:00...	Swaniawski Lowe and Ro...	shopping_pos	317.14	Parks	AZ	35.2563	-111.95	759	Geologist engineering	1957-03-28	3fe3286b...	245324
15	2019-01-01 00:56:00...	Reichert Huels and Hoppe	shopping_net	113.4	Fort Washakie	WY	43.0048	-108.8964	1645	Freight forwarder	1976-02-26	9d660a1...	538148
16	2019-01-01 01:00:00...	Howe Lt	misc_pos	218.71	Littleton	CO	39.5994	-105.0044	320420	Water engineer	1975-07-31	4320404...	846298
17	2019-01-01 01:02:00...	Wolf Inc	grocery_pos	89.11	Meadville	MO	39.7795	-93.3014	964	Tourist information centre manager	1974-12-23	1cb03ad...	616633
18	2019-01-01 01:04:00...	Vandervort-Funk	grocery_pos	50.68	Moab	UT	38.5677	-109.5271	9772	Location manager	1989-11-24	6186137...	561178
19	2019-01-01 01:09:00...	Ledner-Pfannerstill	gas_transport	90.54	Hawthorne	CA	33.9143	-118.3493	93193	Editor magazine features	1995-04-19	295b782f...	26571
20	2019-01-01 01:19:00...	Schaefer McGlynn and Bo...	gas_transport	51.33	Manville	WY	42.73	-104.7024	241	Educational psychologist	1980-12-16	a8af4c63...	607268
21	2019-01-01 01:22:00...	Fisher-Schowalte	shopping_net	226.33	June Lake	CA	37.7773	-119.0825	633	Health service manager	1927-09-09	208e003f...	417687
22	2019-01-01 01:23:00...	Medhurst PLC	shopping_net	215.99	Sixes	OR	42.825	-124.4409	217	Retail merchandiser	1928-10-01	e13482d...	441003
23	2019-01-01 01:23:00...	Kerluke Inc	misc_net	1.47	Holstein	NE	40.4542	-98.6538	331	Telecommunications researcher	1950-08-19	6569a85...	602904
24	2019-01-01 01:23:00...	Bauch-Rayno	grocery_pos	122.05	Westerville	NE	41.4193	-99.3844	73	Product designer	1935-02-10	d83aa73...	438937
25	2019-01-01 01:34:00...	Hills-Olson	grocery_net	27.03	Ballwin	MO	38.577	-90.5255	92608	Engineer communications	2001-07-26	011f7fc2...	281938

After all this steps, data is cleaned.
Now we can get data of users for visualization.

1st Query

We see in this list which categories the customer spends money in different days.

```
SELECT
    customer_ID,
    CONVERT(DATE, trans_datetime) AS trans_date,
    category,
    SUM(amt) AS total_amount
FROM Purchases
GROUP BY customer_ID, CONVERT(DATE, trans_datetime), category
```

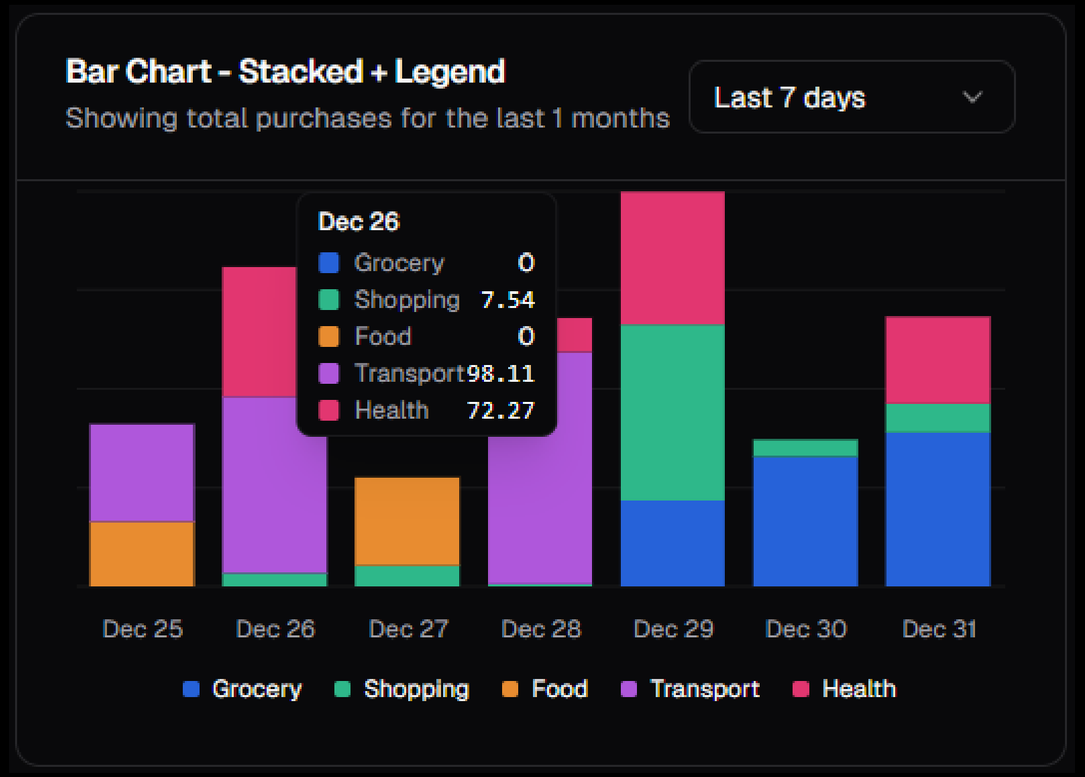
	customer_ID	trans_date	category	total_amount
1	6726	2019-01-03	health_fitness	27.99
2	6726	2019-01-05	shopping_pos	1.84
3	6726	2019-01-06	gas_transport	65.06
4	6726	2019-01-10	grocery_pos	145.86
5	6726	2019-01-10	kids_pets	52.02
6	6726	2019-01-11	kids_pets	129.18
7	6726	2019-01-17	gas_transport	108.41
8	6726	2019-01-18	entertainment	89.47
9	6726	2019-01-18	kids_pets	10.86
10	6726	2019-01-20	health_fitness	150.66
11	6726	2019-01-21	kids_pets	15.75
12	6726	2020-03-07	gas_transport	22.22
13	6726	2020-03-08	gas_transport	18.96
14	6726	2020-03-08	shopping_net	2078.75
15	6726	2020-12-24	health_fitness	173.77
16	6726	2020-12-25	personal_care	28.82
17	6726	2020-12-26	shopping_net	1.13
18	6726	2020-12-27	food_dining	23.28
19	6726	2020-12-29	food_dining	5.73

2nd Query

With the help of Query 2, we can easily see the total money spent on categories as separate columns.

```
WITH CustomerPurchases AS (
    SELECT
        customer_ID, CONVERT(DATE, trans_datetime) AS trans_date, category, amt
    FROM Purchases
)

SELECT
    customer_ID,
    trans_date,
    ISNULL([grocery_net], 0) AS grocery,
    ISNULL([shopping_net], 0) AS shopping,
    ISNULL([food_dining], 0) AS food,
    ISNULL([gas_transport], 0) AS transport,
    ISNULL([health_fitness], 0) AS health
FROM CustomerPurchases
PIVOT (
    SUM(amt)
    FOR category IN ([grocery_net], [shopping_net], [food_dining], [gas_transport],
[health_fitness])
) AS PivotPurchases
ORDER BY customer_ID, trans_date;
```



	customer_ID	trans_date	grocery	shopping	food	transport	health
1	6726	2019-01-03	0	0	0	0	27.99
2	6726	2019-01-05	0	0	0	0	0
3	6726	2019-01-06	0	0	0	65.06	0
4	6726	2019-01-10	0	0	0	0	0
5	6726	2019-01-11	0	0	0	0	0
6	6726	2019-01-17	0	0	0	108.41	0
7	6726	2019-01-18	0	0	0	0	0
8	6726	2019-01-20	0	0	0	0	150.66
9	6726	2019-01-21	0	0	0	0	0
10	6726	2020-03-07	0	0	0	22.22	0
11	6726	2020-03-08	0	2078.75	0	18.96	0
12	6726	2020-12-24	0	0	0	0	173.77
13	6726	2020-12-25	0	0	0	0	0
14	6726	2020-12-26	0	1.13	0	0	0
15	6726	2020-12-27	0	0	23.28	0	0
16	6726	2020-12-29	0	0	5.73	0	0
17	6726	2020-12-30	0	0	0	0	0
18	6726	2020-12-31	0	0	0	0	0
19	8073	2019-01-01	0	0	0	0	159.44
20	8073	2019-01-02	0	0	0	35.27	0
21	8073	2019-01-03	0	0	0	84.68	58.56
22	8073	2019-01-04	0	0	0	0	0
23	8073	2019-01-05	0	1.27	0	78.98	0
24	8073	2019-01-06	0	0	0	0	4.13
25	8073	2019-01-07	0	9	36.78	106.92	99.19

Additional Resources

- Import CSV File to Table -> https://youtu.be/i2cncS_5gIM
- Pivot in SQL -> https://youtu.be/ozy31aJpW-o