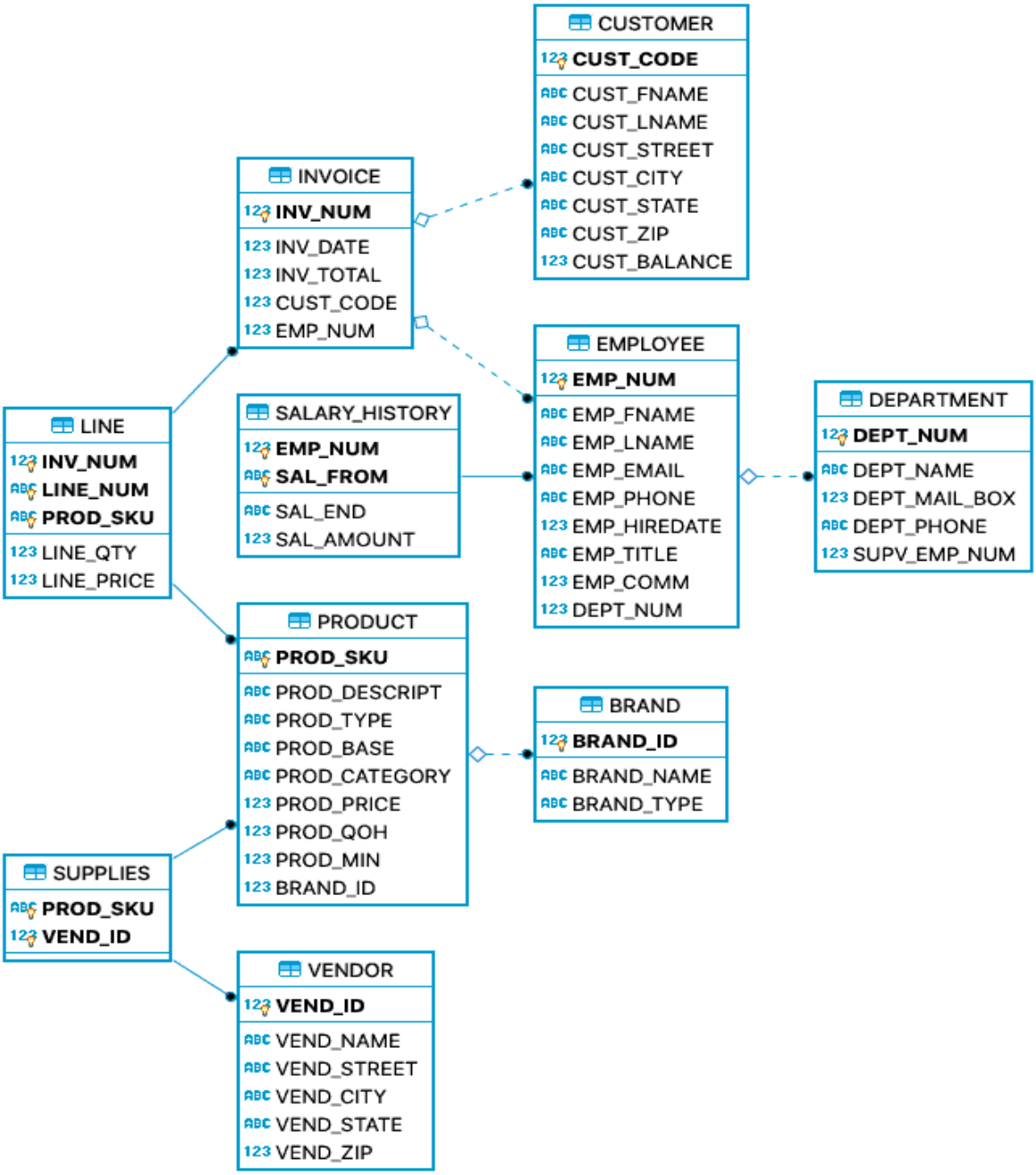
The issue of data cleaning:

When I clean the file, I found above of questions.

- CUST_ZIP has different digits of number. I search the US official website to check the zip-code and then fill digits with 0 to match the number of digits which are 5.
- VEND_STREET value missing
- VEND_NAME value missing
- SAL_END value missing
- There are many unnecessary spaces everywhere. I drop all the spaces for convivence to do query.
- There is same problem in VEND_ID as CUST_ZIP
- There are special marks in PROD_QOH. I drop the value with special number, special marks.
- There are many values zero in EMP_COMM.
- EMP_HIREDATE should be consisted of number with 5 digits.
- There is some situation that two different names in single invoices.
- Change "EMP_ID" to "EMP_NUM" in table INVOICE which can connect table EMPLOEE

ERD

I set the foreign keys in these tables:
- the foreign key "DEPT_NUM" in EMPLOYEE for connecting the DEPARTMENT which primary key is "DEPT_NUM"
- the foreign key "CUST_CODE" and " EMP_NUM" in INVOICE for connecting the CUSTOMER and EMPLOYEE
- the foreign key "INV_NUM" and "PROD_SKU" in LINE for connecting the INVOICE and PRODUCT
- the foreign key "BRAND_ID" in PRODUCT for connecting the BRAND which primary key is "BRAND_ID"
- the foreign key "EMP_NUM" in SALARY_HISTORY for connecting the EMPLOEE which primary key is "EMP_NUM"


CUSTOMER
CUST_CODE
CUST_FNAME
CUST_LNAME
CUST_STREET
CUST_CITY
CUST_STATE
CUST_ZIP
CUST_BALANCE
INVOICE
INV_NUM
INV_DATE
INV_TOTAL
CUST_CODE
EMP_NUM
EMPLOYEE
EMP_NUM
EMP_FNAME
EMP_LNAME
EMP_EMAIL
EMP_PHONE
EMP_HIREDATE
EMP_TITLE
EMP_COMM
DEPT_NUM
DEPARTMENT
DEPT_NUM
DEPT_NAME
DEPT_MAIL_BOX
DEPT_PHONE
SUPV_EMP_NUM
SALARY_HISTORY
EMP_NUM
SAL_FROM
SAL_END
SAL_AMOUNT
LINE
INV_NUM
LINE_NUM
PROD_SKU
LINE_QTY
LINE_PRICE
PRODUCT
PROD_SKU
PROD_DESCRIPT
PROD_TYPE
PROD_BASE
PROD_CATEGORY
PROD_PRICE
PROD_QOH
PROD_MIN
BRAND_ID
BRAND
BRAND_ID
BRAND_NAME
BRAND_TYPE
SUPPLIES
PROD_SKU
VEND_ID
VENDOR
VEND_ID
VEND_NAME
VEND_STREET
VEND_CITY
VEND_STATE
VEND_ZIP

Questions:

1. Write a query to display the current salary for each employee in department 300. Assume that only current employees are kept in the system, and therefore the most current salary for each employee is the entry in the salary history with a NULL end date. Sort the output in descending order by salary amount.

```
SELECT EMP_NUM, SAL_AMOUNT
FROM SALARY_HISTORY
WHERE SAL_END = ' - '
AND EMP_NUM in (
SELECT EMP_NUM
FROM EMPLOYEE
WHERE DEPT_NUM = 300
)
ORDER BY SAL_AMOUNT DESC
```

	EMP_NUM	SAL_AMOUNT
1	83,746	95,550
2	84,328	94,090
3	83,716	85,920
4	84,432	85,360
5	83,902	79,540
6	83,695	79,200
7	84,500	78,690
8	84,594	77,400
9	83,910	76,110
10	83,359	72,240
11	83,790	72,000
12	83,433	68,870
13	84,521	66,000
14	83,653	61,920
15	83,738	58,200
16	83,788	56,760
17	83,867	56,750
18	84,234	54,720
19	83,637	52,870
20	83,877	52,650
21	84,035	51,600
22	83,729	48,500
23	83,732	44,720
24	83,644	43,200
25	83,312	42,400

2. Write a query to display the starting salary for each employee. The starting salary would be the entry in the salary history with the oldest salary start date for each employee. Sort the output by employee number.

```
SELECT E.EMP_NUM, MIN(SAL_AMOUNT), MIN(SAL_FROM)
FROM EMPLOYEE E JOIN SALARY_HISTORY S
GROUP BY E.EMP_NUM
ORDER BY E.EMP_NUM
```

	EMP_NUM	MIN(SAL_AMOUNT)	MIN(SAL_FROM)
1	83,304	6,190	1978-01-13
2	83,308	6,190	1978-01-13
3	83,312	6,190	1978-01-13
4	83,314	6,190	1978-01-13
5	83,318	6,190	1978-01-13
6	83,321	6,190	1978-01-13
7	83,332	6,190	1978-01-13
8	83,341	6,190	1978-01-13
9	83,347	6,190	1978-01-13
10	83,349	6,190	1978-01-13
11	83,359	6,190	1978-01-13
12	83,366	6,190	1978-01-13
13	83,371	6,190	1978-01-13
14	83,372	6,190	1978-01-13
15	83,374	6,190	1978-01-13
16	83,378	6,190	1978-01-13
17	83,382	6,190	1978-01-13
18	83,385	6,190	1978-01-13
19	83,398	6,190	1978-01-13
20	83,403	6,190	1978-01-13
21	83,404	6,190	1978-01-13
22	83,411	6,190	1978-01-13
23	83,413	6,190	1978-01-13
24	83,415	6,190	1978-01-13
25	83,419	6,190	1978-01-13
26	83,423	6,190	1978-01-13
27	83,428	6,190	1978-01-13
28	83,432	6,190	1978-01-13
29	83,433	6,190	1978-01-13
30	83,434	6,190	1978-01-13
31	83,437	6,190	1978-01-13

There are total 363 rows and I screenshot 31 rows of them.

3. Write a query to display the invoice number, line numbers, product SKUs, product descriptions, and brand ID for sales of sealer and top coat products of the same brand on the same invoice.

```
SELECT b.INV_NUM, b.LINE_NUM, a.PROD_SKU, a.PROD_DESCRIPT, a.BRAND_ID
FROM PRODUCT a, LINE b
WHERE a.PROD_SKU = b.PROD_SKU
and a.PROD_CATEGORY in ('Top Coat', 'Sealer')
ORDER BY 1,2,3
```

	INV_NUM	LINE_NUM	PROD_SKU	PROD_DESCRIPT	BRAND_ID
1	401	3	1200-KBU	Stain, Semi-Transparent, for Interior Wood	25
2	436	4	9288-IRF	Varnish, Water Based, Clear (eggshell-like) MPI Glo	25
3	506	6	2046-RUS	Polyurethane Deck Coating (Slip-Resistant)	28
4	515	5	5529-SBL	Water Repellent, Clear (Paintable)	30
5	560	2	1663-CDD	Light Industrial Coating, Exterior, Water Based, Sem	25
6	665	5	9569-WMK	Epoxy, High Build, Low Gloss	29
7	669	5	9671-NYZ	Lacquer, Pigmented, Gloss	30
8	672	2	2262-NBI	Vinyl, High Build	31
9	685	6	3701-YAW	Sealer, Solvent Based, for Concrete Floors	30
10	696	1	8726-ZNM	Floor Paint, Alkyd, Low Gloss	29
11	740	3	7006-IXN	Polyurethane, Moisture Cured, Pigmented, Intermed	29
12	745	6	6109-ZZO	Deck Coating, Latex, Exterior	23
13	757	2	6488-MIG	Latex, Recycled (Consolidated), Interior (MPI Gloss	30
14	784	4	2366-EFV	Varnish, Interior, Polyurethane, Oil Modified, Gloss	30
15	793	2	6894-JQV	Latex, Recycled (Consolidated), Interior (MPI Gloss	27
16	825	2	5140-RTG	Fire Resistant Sealer, for Exterior Wood (ULC Appro	35
17	840	6	6858-EJW	Alkyd, Exterior Gloss (MPI Gloss Level 6)	30
18	844	5	1045-DUY	Traffic Marking Paint, Alkyd	35
19	855	3	7532-PYJ	Floor Paint, Latex, Gloss	25
20	859	6	5653-RTU	Aluminum Paint	23
21	883	6	5379-BLX	Dry Fall, Water Based, for Galvanized Steel, MPI Glo	28
22	910	6	5659-BFS	Elastomeric Coating, Exterior, Water Based, Non-Fla	35
23	912	2	2068-TSC	Graffiti Protection Top-Coat	30
24	915	1	8726-ZNM	Floor Paint, Alkyd, Low Gloss	29
25	926	6	9288-IRF	Varnish, Water Based, Clear (eggshell-like) MPI Glo	25
26	944	4	1198-STR	Latex, Exterior, Gloss (MPI Gloss Level 6)	28
27	946	3	7224-FEU	Latex, Recycled (Consolidated), Exterior, Low Sheer	27
28	977	5	1045-DUY	Traffic Marking Paint, Alkyd	35
29	982	1	2584-CIJ	Stain, for Exterior Wood Decks	30
30	994	5	7158-LJP	Textured Coating, Latex, Flat	33
31	1,000	3	5213-YTB	Latex, Interior, High Performance Architectural, ('ve	31

There are total 1015 rows and I screenshot 31 rows of them.

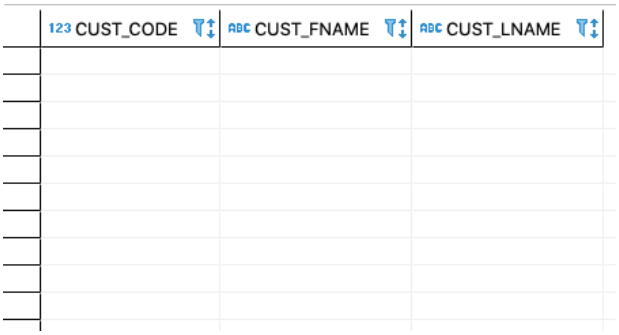
4. The Binder Prime Company wants to recognize the employee who sold the most of their products during a specified period. Write a query to display the employee number, employee first name, employee last name, e-mail address, and total units sold for the employee who sold the most Binder Prime brand products between November 1, 2015, and December 5, 2015. If there is a tie for most units sold, sort the output by employee last name.

```
select e.EMP_NUM,e.EMP_FNAME,e.EMP_LNAME,e.EMP_EMAIL,count(i.INV_NUM)
as units_sold
from EMPLOYEE e inner join INVOICE I on e.EMP_NUM=i.EMP_NUM inner join
LINE l on l.INV_NUM=i.INV_NUM inner join
PRODUCT p on p.PROD_SKU=l.PROD_SKU INNER join BRAND b on
b.BRAND_ID=p.BRAND_ID
WHERE b.BRAND_NAME = 'BINDER PRIME' group by
e.EMP_NUM,e.EMP_FNAME,e.EMP_LNAME,e.EMP_EMAIL
and i.INV_DATE BETWEEN '01-NOV-15' AND '05-DEC-15' order by
units_sold, e.EMP_LNAME
```

	EMP_NUM	EMP_FNAME	EMP_LNAME	EMP_EMAIL	units_sold
1	83,537	CLEO	ENGLISH	E.CLEO90@LGCOMPANY.COM	1
2	83,649	DELMA	JACOB	J.DELMA91@LGCOMPANY.COM	1
3	83,878	PAT	MARIN	M.PAT95@LGCOMPANY.COM	1
4	84,364	LESLIE	VARNER	V.LESLIE0@LGCOMPANY.COM	1
5	84,276	ROSALIND	VILLARREAL	V.ROSALI99@LGCOMPANY.COM	1
6	83,792	WALLY	ANDERSEN	A.WALLY94@LGCOMPANY.COM	2
7	83,936	BRADFORD	BRAY	B.BRADFO96@LGCOMPANY.COM	2
8	83,763	JAIME	FELTON	F.JAIME94@LGCOMPANY.COM	2
9	83,820	TOM	HOBSON	H.TOM94@LGCOMPANY.COM	2
10	83,734	INEZ	ROCHA	R.INEZ93@LGCOMPANY.COM	2
11	84,248	DANICA	CASTLE	C.DANICA99@LGCOMPANY.COM	3
12	84,420	DOUG	CAUDILL	C.DOUG0@LGCOMPANY.COM	3
13	84,021	JAROD	DICKINSON	D.JAROD97@LGCOMPANY.COM	3
14	84,306	ROWENA	MEDINA	M.ROWENA99@LGCOMPANY.COM	3
15	83,593	ROSANNE	NASH	N.ROSANN91@LGCOMPANY.COM	3
16	83,509	FRANKLYN	STOVER	S.FRANKL89@LGCOMPANY.COM	3
17	83,621	FONDA	GONZALEZ	G.FONDA91@LGCOMPANY.COM	4
18	84,191	ROXANA	HOLBROOK	H.ROXANA98@LGCOMPANY.COM	4
19	83,705	JOSE	BARR	B.JOSE92@LGCOMPANY.COM	5
20	83,993	SANG	CORTES	C.SANG97@LGCOMPANY.COM	5
21	84,163	GWEN	EASLEY	E.GWEN98@LGCOMPANY.COM	5
22	84,334	LINNIE	GOLDMAN	G.LINNIE99@LGCOMPANY.COM	5
23	83,964	HAILEY	SWEENEY	S.HAILEY97@LGCOMPANY.COM	5
24	83,565	LOURDES	ABERNATHY	A.LOURDE90@LGCOMPANY.COM	6
25	84,049	LANE	BRANDON	B.LANE97@LGCOMPANY.COM	6
26	83,906	DELLA	SIMONS	S.DELLA96@LGCOMPANY.COM	6
27	84,134	ROSALIE	GARLAND	G.ROSALI98@LGCOMPANY.COM	7
28	83,677	HERB	MANNING	M.HERB92@LGCOMPANY.COM	7
29	84,106	FELICE	SAMUEL	S.FELICE98@LGCOMPANY.COM	7
30	84,219	THURMAN	WILKINSON	W.THURMA99@LGCOMPANY.COM	8
31	84,078	DIEGO	ERWIN	E.DIEGO98@LGCOMPANY.COM	9
32	83,850	RUSTY	MILES	M.RUSTY95@LGCOMPANY.COM	9

5. Write a query to display the customer code, first name, and last name of all customers who have had at least one invoice completed by employee 83649 and at least one invoice completed by employee 83677. Sort the output by customer last name and then first name.

```
SELECT i.CUST_CODE, c.CUST_FNAME, c.CUST_LNAME
FROM INVOICE AS i
INNER JOIN CUSTOMER AS c
ON i.CUST_CODE = c.CUST_CODE
WHERE i.CUST_CODE in
(SELECT i.CUST_CODE
FROM INVOICE AS i
GROUP BY i.CUST_CODE
HAVING count(i.INV_NUM) > 1)
AND i.INV_NUM in
(SELECT i.INV_NUM
FROM INVOICE AS i
WHERE i.EMP_NUM = "83649")
AND i.INV_NUM in
(SELECT i.INV_NUM
FROM INVOICE AS i
WHERE i.EMP_NUM = "83677")
ORDER BY c.CUST_FNAME, c.CUST_LNAME;
```



6. LargeCo is planning a new promotion in Alabama (AL) and wants to know about the largest purchases made by customers in that state. Write a query to display the customer code, customer first name, last name, full address, invoice date, and invoice total of the largest purchase made by each customer in Alabama. Be certain to include any customers in Alabama who have never made a purchase (their invoice dates should be NULL and the invoice totals should display as 0).

```
SELECT c.CUST_CODE, c.CUST_FNAME, c.CUST_LNAME,
CONCAT(c.CUST_STREET,', ', c.CUST_CITY,', ', c.CUST_STATE,' ',
c.CUST_ZIP) AS CUST_ADDRESS,
i.INV_DATE, COUNT(i.INV_TOTAL) AS LARGEST_INV_TOTAL
FROM customer AS c
LEFT JOIN invoice AS i
ON c.CUST_CODE = i.CUST_CODE
WHERE c.CUST_STATE = 'AL'
GROUP BY c.CUST_CODE, c.CUST_FNAME, c.CUST_LNAME,
CUST_ADDRESS,i.INV_DATE
order by LARGEST_INV_TOTAL;
```

	CUST_CODE	CUST_FNAME	CUST_LNAME	CUST_ADDRESS	INV_DATE	LARGEST_INV_TOTAL
18	1,264	MICHAELA	RICHARD	44 RASMUSSON STREET, Georgetown, AL 3652	12	1
19	1,407	FELICIA	CRUZ	643 TURNAGAIN PARKWAY, Coalburg, AL 3506	1	1
20	1,443	ALYSON	SELF	772 LUPIN DRIVE, Motley, AL 36276	1	1
21	380	ALBINA	ENGLE	670 UPPER BOWERY LANE, Clanton, AL 35045	4	1
22	738	ALIDA	HANSEN	792 FERGY CIRCLE, Furman, AL 36741	8	1
23	979	IMOGENE	MAYES	1017 HARCA STREET, Sylacauga, AL 35150	9	1
24	169	ROSS	LANG	1991 EASTWIND COURT, Higdon, AL 35979	11	1
25	1,275	ELEANORE	NEFF	556 CACHE DRIVE, Saint Stephens, AL 36569	12	1
26	780	LARISSA	POOL	574 ADAK CIRCLE, Decatur, AL 35609	1	1
27	538	CHIQUITA	CALDWELL	1501 BRIGGS COURT, Normal, AL 35762	5	1
28	820	MARCELA	DUGAN	1785 DORIS PLACE, Sylacauga, AL 35150	8	1
29	1,029	JOHNETTA	ROY	1163 GIROUX CIRCLE, Dauphin Island, AL 3652	9	1
30	152	LISETTE	WHITTAKER	339 NORTHPARK DRIVE, Montgomery, AL 3619	11	1
31	925	ALANA	BOOKER	1874 I STREET, Mccullough, AL 36502	12	1
32	1,412	EVALYN	HEWITT	293 TIMOTHY CIRCLE, Tallassee, AL 36078	1	1
33	643	NINA	ALLEN	680 RED TALON DRIVE, Robertsdale, AL 36574	6	1
34	188	LUANNE	GOODWIN	293 KIANA AVENUE, Pinegrove, AL 36507	8	1
35	1,100	ELEANORE	SAUNDERS	820 QUARTZ AVENUE, Silver Cross, AL 36538	10	1
36	219	CATHI	WHITEHEAD	760 WOODCLIFF DRIVE, Huntsville, AL 35893	11	1
37	1,172	ADELE	PERKINS	1192 RICHARDSON VISTA ROAD, Sylacauga, AL	12	1
38	1,248	LISA	BRADY	491 LOWLAND AVENUE, Daphne, AL 36577	1	1
39	696	ALISHA	TOMLINSON	1985 EAST 52ND AVENUE, Catherine, AL 3672	7	1
40	855	AUBREY	GLOVER	907 GOLD CLAIM DRIVE, Honoraville, AL 36042	8	1
41	1,068	ELIZA	CURRIE	778 LOUDERMILK CIRCLE, Panola, AL 35477	10	1
42	1,233	NATHALIE	CHURCH	1802 SNOWY OWL CIRCLE, Napier Field, AL 36	11	1
43	458	ELOISA	VALLE	182 BRANDON STREET, Abel, AL 36258	12	1
44	89	MONICA	CANTRELL	697 ADAK CIRCLE, Loachapoka, AL 36865	1	1
45	364	DELLA	MAYO	543 STELIOS CIRCLE, Birmingham, AL 35214	7	1
46	886	ROSARIO	STOKES	959 SUNRISE DRIVE, Hightower, AL 36263	8	1
47	903	ROBIN	ADDISON	323 LORETTA PLACE, Mobile, AL 36693	10	1
48	853	GAYLORD	BOLTON	1069 LUGENE LANE, Montgomery, AL 36131	11	1
49	1,350	LATONYA	KAY	61 LOUSSAC DRIVE, Seaboard, AL 36529	12	1

There are total 50 rows and I screenshot 49 rows of them.

7. One of the purchasing managers is interested in the impact of product prices on the sale of products of each brand. Write a query to display the brand name, brand type, average price of products of each brand, and total units sold of products of each brand. Even if a product has been sold more than once, its price should only be included once in the calculation of the average price. However, you must be careful because multiple products of the same brand can have the same price, and each of those products must be included in the calculation of the brand's average price.

```
SELECT f1.brand_name, f1.brand_type, round(f1.average_price,2) AS
AvgPrice, f2.amt_sold AS UnitsSold FROM
( SELECT b.brand_name, b.brand_type, b.brand_id,
(sum(p.prod_price)/count(distinct p.prod_sku)) AS average_price
FROM product AS p, brand AS b
WHERE b.brand_id=p.brand_id
GROUP BY b.brand_id
ORDER BY b.brand_name
) AS f1
INNER JOIN
( SELECT b.brand_id, sum(l.line_qty) AS amt_sold
FROM brAND AS b, line AS l, product AS p
WHERE b.brand_id = p.brand_id
AND p.prod_sku = l.prod_sku
GROUP BY b.brand_id
ORDER BY b.brand_name
) AS f2
ON f1.brand_id = f2.brand_id
ORDER BY round(f1.average_price,2) desc;
```

	brand_name	brand_type	AvgPrice	UnitsSold
1	BUSTERS	VALUE	22.59	479
2	HOME COMFORT	CONTRACTOR	21.80	466
3	FORESTERS BEST	VALUE	20.94	221
4	LONG HAUL	CONTRACTOR	20.12	665
5	LE MODE	PREMIUM	19.22	561
6	OLDE TYME QUALITY	CONTRACTOR	18.33	417
7	VALU-MATTE	VALUE	16.84	312
8	STUTTENFURST	CONTRACTOR	16.47	401
9	BINDER PRIME	PREMIUM	16.12	387

8. The purchasing manager is still concerned about the impact of price on sales. Write a query to display the brand name, brand type, product SKU, product description, and price of any products that are not a premium brand, but that cost more than the most expensive premium brand products.

```
SELECT b.BRAND_NAME,b.BRAND_TYPE, p.PROD_SKU, p.PROD_DESCRIPT,
p.PROD_PRICE
FROM PRODUCT AS p
INNER JOIN BRAND AS b
ON p.BRAND_ID = b.BRAND_ID
WHERE b.BRAND_TYPE != "premium"
AND p.PROD_PRICE >
(SELECT max(p.PROD_PRICE)
FROM product AS p
INNER JOIN BRAND AS b
ON p.BRAND_ID = b.BRAND_ID
WHERE b.BRAND_TYPE = "premium");
```

	BRAND_NAME	BRAND_TYPE	PROD_SKU	PROD_DESCRIPT	PROD_PRICE
1	LONG HAUL	CONTRACTOR	1964-OUT	Fire Resistant Top Coat, for Interior Wood	78.4900

9. Using SQL descriptive statistics functions calculate the value of the following items:
What are the products that have a price greater than $50?

```
SELECT *
FROM PRODUCT AS p
WHERE p.PROD_PRICE > 50
```

	PROD_SKU	PROD_DESCRIPT	PROD_TYPE	PROD_BASE	PROD_CAT	PROD_PR	PROD_QO	PROD_MIN	BRAND_ID
1	1021-MTI	Elastomeric, Exterior, Industrial Grade, Water Bas	Exterior	Water	Top Coat	62.9900	22	25	35
2	1964-OUT	Fire Resistant Top Coat, for Interior Wood	Interior	Solvent	Top Coat	78.4900	120	10	30
3	3694-XFJ	Epoxy-Modified Latex, Interior, Semi-Gloss (MPI	Interior	Water	Top Coat	54.8900	39	25	27

What is total value of our entire inventory on hand?

```
SELECT sum(p.PROD_PRICE*p.PROD_QOH) AS TotalValueOfInventory
FROM PRODUCT AS p
```

	TotalValueOfInventory
1	360,307.7926878929

How many customers do we presently have and what is the total of all customer balances?

```
SELECT count(cust_code) AS NumberOfCustomers, sum(cust_balance) AS
TotalAllCustomerBalance
FROM CUSTOMER;
```

	NumberOfCustomers	TotalAllCustomerBalance
1	1,362	787,201.150834322

What are to top three states that buy the most product in dollars from the company?

```
SELECT c.CUST_STATE, sum(i.INV_TOTAL) AS AmountSpendTotal
FROM CUSTOMER AS c
INNER JOIN INVOICE AS i
ON c.CUST_CODE = i.CUST_CODE
GROUP BY c.CUST_STATE
ORDER BY sum(i.INV_TOTAL) DESC;
```

	CUST_STATE	AmountSpendTotal
1	PA	37,896.1098718643
2	NY	31,974.7099123001
3	NC	19,311.0700931549
4	OH	18,767.4100437164
5	IN	16,668.0199565887
6	FL	16,186.0600302219
7	MD	13,030.1899995804
8	KY	12,532.3400235176
9	MA	11,592.9299736023
10	VA	11,311.8799552917
11	AL	10,118.6100215912
12	VT	8,725.4899463654
13	NJ	8,454.9299697876
14	MI	8,039.2200527191
15	WV	7,949.9499435425
16	GA	7,772.1699810028
17	MS	7,421.629983902
18	TN	7,388.5300254822
19	SC	6,567.760011673
20	ME	4,176.2100315094
21	CT	3,868.0899915695
22	NH	2,396.6200218201
23	RI	1,384.9299888611
24	DE	1,087.549987793

10. Using predictive statistics calculate what the predicted forecast of sales for the next year based on the INV_DATE (independent) and INV_TOTAL (dependent).

If I use the variable INV_DATE, which is date, to explain, it would not be allowed to explain the dependent variable INV_TOTAL. The type of the INV_TOTAL is numerical. If I do the regression, the independent variable should be the same type with INV_TOTAL. I clean the data to eliminate the duplicated date, which means, there are many INV_TOTAL in the same date. I calculate their mean, assigning the number as the INV_TOTAL to the specific day.

The regression model doesn't fit. I try to use the autoregression model to explain, which is using the lagging value of INV_TOTAL to be explanatory variable and do the analysis. The following I will show the model:

$$INV\_TOTAL_t = \beta_0 + \beta_1 INV\_TOTAL_{t-1} + \beta_2 INV\_TOTAL_{t-2} + e_t$$

Then to use two lagging values establishes the autoregressive model. The following would show the result of the regression:

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.02764							
R Square	0.000764							
Adjusted R Square	-0.006922							
Standard Error	74.195659							
Observations	263							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	2	1094.3058	547.15291	0.0993921	0.9054221			
Residual	260	1431298.9	5504.9958					
Total	262	1432393.2						
	Coefficients	Standard Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	207.11791	18.197806	11.381477	1.27E-24	171.28406	242.95176	171.28406	242.95176
lag1	-0.014111	0.0620169	-0.227538	0.8201844	-0.136231	0.1080082	-0.136231	0.1080082
lag2	-0.023963	0.062012	-0.386426	0.6994971	-0.146073	0.0981467	-0.146073	0.0981467

$$INV\_TOTAL_t = 207.12 - 0.014\ INV\_TOTAL_{t-1} + -0.024\ INV\_TOTAL_{t-2}$$

From the figure, the coefficients of the two lagging are around -0.014 and - 0.024, which means the INV_TOTAL has negative relationship with its lagging value. However, the p-value of them are larger than 0.05, which tells us those are not statistically significant. The INV_TOTAL may not have significant relationship with its lagging value.

Additionally, the R^2 is about 0.07%, which means there are only 0.07% of the INV_TOTAL can be explained by its lag values. I think the percentage is too small to explain.