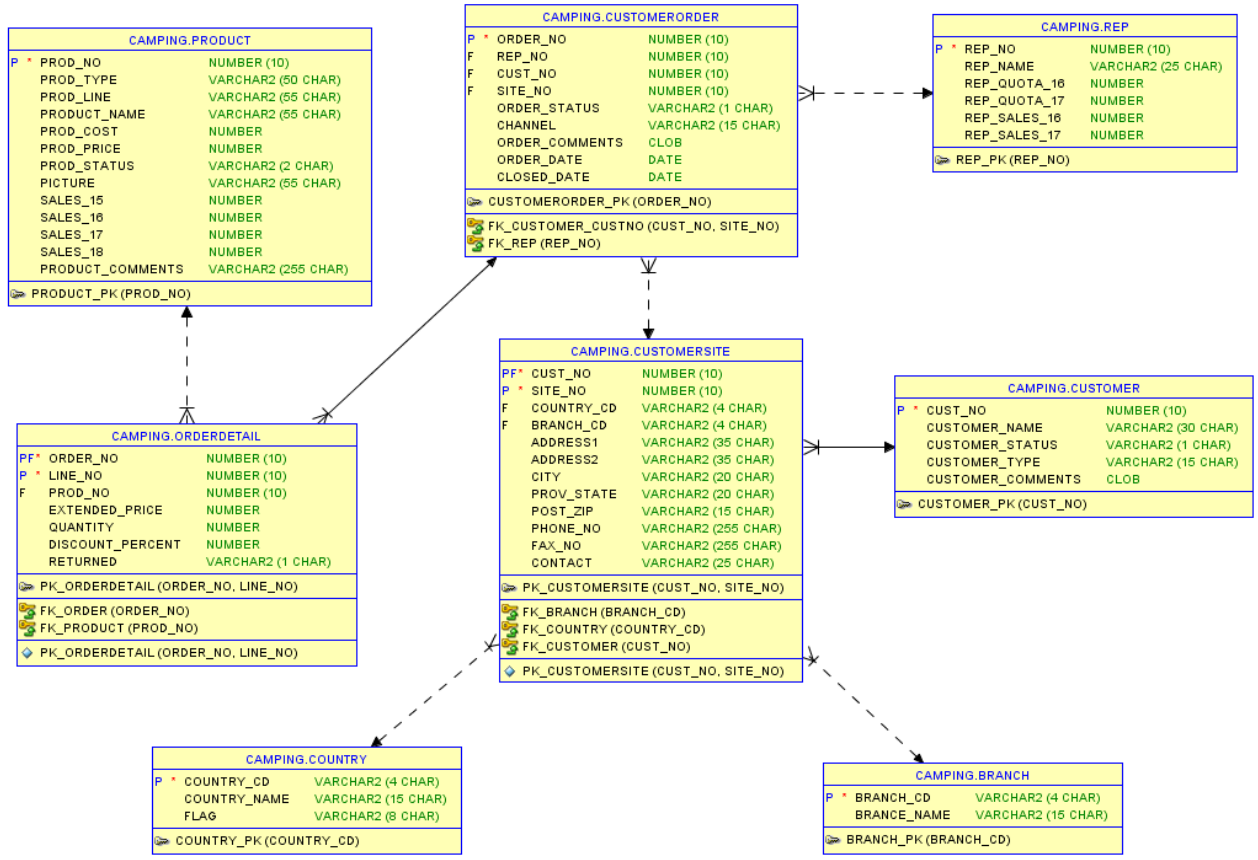
Question 1. Reverse engineer the Camping Store schema into a physical level diagram.



Question 2. Show the total dollar amount sold to customers summarized by state/province of their customer site address and each month of each Year (YYYY-MM).

- SQL:

```
SELECT  TO_CHAR (o.order_date, 'YYYY-MM') AS Order_Month,
        NVL (s.prov_state, s.city) AS State_Province,
        SUM (od.quantity * od.extended_price) AS Total_Dollars
FROM   (camping.orderdetail od INNER JOIN camping.customerorder o
              ON od.order_no = o.order_no)
        INNER JOIN camping.customersite s ON o.cust_no = s.cust_no
                 AND o.site_no = s.site_no
WHERE          od.returned = 'N'
GROUP BY       TO_CHAR (o.order_date, 'YYYY-MM'),
               NVL (s.prov_state, s.city)
ORDER BY       TO_CHAR (o.order_date, 'YYYY-MM');
```

- Results: (first 24 of 161 rows)

	ORDER_MONTH	STATE_PROVINCE	TOTAL_DOLLARS
1	2018-01	California	9305
2	2018-01	Monterey	10432
3	2018-02	California	11024
4	2018-02	Cork	14778
5	2018-02	Frankfurt	18104
6	2018-02	Hong Kong	11804
7	2018-02	NSW	11090
8	2018-02	Nova Scotia	9154
9	2018-02	Quebec	10068
10	2018-02	Sevilla	16221
11	2018-03	Brighton	13150
12	2018-03	Goteburg	15728
13	2018-03	London	24185
14	2018-03	Monterey	3005
15	2018-03	Quebec	10754
16	2018-03	Stockholm	7035
17	2018-03	Texas	8565
18	2018-03	Tokyo	9597
19	2018-04	Berkshire	10213
20	2018-04	California	30633
21	2018-04	Colorado	12523
22	2018-04	Hong Kong	15924
23	2018-04	Paris	14425
24	2018-04	Stockholm	16871

Question 3. Show the total dollar amount sold summarized by product name and each month of each Year (YYYY-MM) along with RANK (order by increasing RANK). Only include sales by reps who met or exceeded their 2017 sales quota.

- SQL:

```
SELECT  RANK() OVER (ORDER BY SUM (od.quantity * od.extended_price) DESC)
              AS Total_Dollar_Rank,
        TO_CHAR (o.order_date, 'YYYY-MM') AS Order_Month,
        p.product_name,
        SUM (od.quantity * od.extended_price) AS Total_Dollars
FROM ((camping.product p INNER JOIN camping.orderdetail od
                ON p.prod_no = od.prod_no)
     INNER JOIN camping.customerorder o ON od.order_no = o.order_no)
     INNER JOIN camping.rep r ON o.rep_no = r.rep_no
WHERE      od.returned = 'N'
AND        r.rep_sales_17 >= r.rep_quota_17
GROUP BY   TO_CHAR (o.order_date, 'YYYY-MM'), p.product_name;
```

- Result: (first 24 of 277 rows)

	TOTAL_DOLLAR_RANK	ORDER_MONTH	PRODUCT_NAME	TOTAL_DOLLARS
1	1	2019-03	Star Lite	63304
2	2	2019-09	Pocket Water Filter	42607
3	2	2019-10	Pocket Water Filter	42607
4	4	2019-10	Star Lite	35126
5	5	2019-12	Pocket Water Filter	27057
6	6	2019-06	Pocket Water Filter	18660
7	7	2019-08	Pocket Water Filter	17727
8	8	2019-12	Star Gazer-2	13048
9	9	2019-10	Sun Shelter-30	12826
10	10	2019-03	Pocket Radon Alerter	11868
11	11	2019-11	GO Duffel Bag	9920
12	12	2019-01	Star Gazer-3	9500
13	13	2019-04	Pocket Water Filter	9330
14	14	2019-01	RiverKind Shampoo	9284
15	15	2019-12	StarDome	8864
16	16	2018-02	Pocket Water Filter	8370
17	17	2019-03	Sun Shelter-15	7667
18	18	2019-02	RiverKind Shampoo	7062
19	19	2019-07	RiverKind Shampoo	6655
20	20	2018-02	Star Lite	6600
21	21	2019-06	RiverKind Soap	6110
22	22	2019-06	GO Duffel Bag	5952
23	23	2019-07	Sun Shelter-15	5753
24	24	2019-03	Sun Shelter-30	5709

Question 4. Write a SQL statement to show all Canadian based customer's name, state/province and total dollar amount sold on "Outdoor Products" or "Environmental Line" product types.

- SQL:

```
SELECT      c.customer_name AS Customer_Name,
            s.prov_state AS State_Province,
            SUM (od.quantity * od.extended_price) AS Total_Dollars
FROM    ((((camping.customer c INNER JOIN camping.customersite s
                ON c.cust_no = s.cust_no)
      INNER JOIN camping.country ct ON s.country_cd = ct.country_cd)
      INNER JOIN camping.customerorder o ON s.cust_no = o.cust_no
                                        AND s.site_no = o. site_no)
       INNER JOIN camping.orderdetail od ON o.order_no = od.order_no)
      INNER JOIN camping.product p ON od.prod_no = p.prod_no
WHERE     p.prod_type IN ('Outdoor Products', 'Environmental Line')
AND        od.returned = 'N'
AND        ct.country_name = 'Canada'
GROUP BY     c.customer_name, s.prov_state;
```

- Result:

	CUSTOMER_NAME	STATE_PROVINCE	TOTAL_DOLLARS
1	Laperier Sportifs 5	Quebec	5565
2	Clear Valley Waters 3	Ontario	19072
3	GO Outlet Montreal	Quebec	21882
4	Clear Valley Waters 2	Nova Scotia	15229
5	New Wave Wilderness 2	Ontario	12803
6	Supras Camping Supplies 3	Quebec	11150
7	Kay Mart 4	Quebec	5588
8	Andes Camping Supplies 1	Manitoba	21715
9	Supras Camping Supplies 5	Ontario	31330
10	Excellence en Montagne	Quebec	11732
11	Vacation Central 4	British Columbia	14762
12	Andes Camping Supplies 3	British Columbia	9714

Question 5. Show the product name and total profit for the product that has the largest profit margin (extended price compared to product cost) when sold over the internet.

- SQL:

```
SELECT p.product_name,
       ROUND(SUM((od.extended_price - p.prod_cost)/od.extended_price),2)
            AS Profit_Margin
FROM (camping.product p INNER JOIN camping.orderdetail od
                ON p.prod_no = od.prod_no)
            INNER JOIN camping.customerorder o
                ON od.order_no = o.order_no
WHERE       o.channel = 'Internet Sales'
AND         od.returned = 'N'
GROUP BY    p.product_name
ORDER BY    2 DESC
FETCH FIRST 1 ROW ONLY;
```

- Result:

	PRODUCT_NAME	PROFIT_MARGIN
1	Sun Shelter-30	6,48

Question 6. Who is the "best" customer? Justify your rationale and back it up with queries and data. You may also wish to graph various data to support your justification.

- SQL:

```
SELECT     c.cust_no, c.customer_name,
           SUM(od.extended_price * od.quantity) AS Total_Sales,
           COUNT(o.cust_no) AS Order_Count,
           (SUM(od.extended_price * od.quantity)/COUNT(o.cust_no))
           AS Dollar_Per_Order
FROM camping.customer c INNER JOIN camping.customersite s
            ON c.cust_no = s.cust_no
        INNER JOIN camping.customerorder o
            ON s.cust_no = o.cust_no
        INNER JOIN camping.orderdetail od
            ON o.order_no = od.order_no
WHERE      od.returned = 'N'
GROUP BY   c.cust_no, c.customer_name, o.cust_no
ORDER BY   Order_Count DESC
--ORDER BY Total_Sales DESC
FETCH FIRST 5 ROWS ONLY;
```

- Result 1 (sort by descending Total_Sales):

	CUST_NO	CUSTOMER_NAME	TOTAL_SALES	ORDER_COUNT
1	1102	Pro Form Supplies 4	72976	31
2	1130	Fresh Air Lte 4	61332	36
3	1120	GO Outlet London	59231	23
4	1062	123 Fitness PTE Ltd	51577	23
5	1041	Vacation Central 3	50137	30

- Result 2 (sort by descending Order_Count

	CUST_NO	CUSTOMER_NAME	TOTAL_SALES	ORDER_COUNT
1	1056	OutBack Pty 2	48193	36
2	1130	Fresh Air Lte 4	61332	36
3	1102	Pro Form Supplies 4	72976	31
4	1041	Vacation Central 3	50137	30
5	1048	Supras Camping Supplies 5	36242	24

- Justification:

  Regarding the best customer, the following requirements are expected:

  - The time that customer joins along with the current dates (as retention rates)
  - Whether the customer is active or not
  - The sales from the customer
  - The buying frequency of the customer
  - The frequency of returning items from the customer
  - The customer lifetime value (CLV): $Margin \cdot \frac{Retention\ Rate}{1+Discount\ Rate-Retention\ Rate}$
  - The purchase (sales) velocity of the customer

  However, when looking at the dataset, there are no records of the retention rates (only ordered date,) and all customer status are active; as a result, only `Total_Sales` and `Order_Count` are taken into consideration, along with the `Customer_Name` and their numbers respectively. Since there are two fields, there are two `ORDER BY` statements , one of which should be removed and/or commented when running the SQL script twice. Also, there is no retention rates available, it is unable to compute the CLV as well as purchase velocity (no recoards about the next order in terms of time comparing to the previous order, or the dollar amount between the two adjacent orders…). As a result, based on the result acquired above, I would say the best customer title is awarded to Pro Form Supplies 4 with the customer number of 1102, due to the fact that they have the highest sales of $72,976, which is more $10,000 than Fresh Air Lte 4 – the second highest. Regarding the number of orders, Pro Form Supplies 4 is also the second – highest with 31 orders in total, only surpassed by the 36 total orders from Fresh Air Lte 4 and OutBack Pty 2.

Question 7. Which month (be sure to say from which year) had the largest percentage increase in sales over the prior month? Justify your rationale with SQL query (Hint: Look at the LAG function).

- SQL:

```
SELECT Order_Month,
      ROUND((Order_Sales-Order_Prior_Sales)/(Order_Prior_Sales),2)
                  AS Sales_Percentage_Change
FROM ( SELECT TO_CHAR(o.order_date, 'YYYY-MM') AS Order_Month,
              SUM(od.extended_price*od.quantity) AS Order_Sales,
              LAG (SUM(od.extended_price*od.quantity), 1,
                      SUM(od.extended_price*od.quantity))
                OVER(ORDER BY TO_CHAR(o.order_date, 'YYYY-MM'))
                      AS Order_Prior_Sales
        FROM camping.orderdetail od
                  INNER JOIN camping.customerorder o
                      ON od.order_no = o.order_no
        GROUP BY  TO_CHAR(o.order_date, 'YYYY-MM')
        ORDER BY  TO_CHAR(order_date, 'YYYY-MM')
       )
WHERE      Order_Prior_Sales > 0
AND        od.returned = 'N'
ORDER BY   2 DESC
FETCH FIRST 1 ROW ONLY;
```

- Result:

	ORDER_MONTH	SALES_PERCENTAGE_CHANGE
1	2018-02	4,39

- Justification:

Total sales equals to the products' extended prices time their quantities. Because there are no record of the prior month before January 2018 (December 2017), it is assumed that December 2017's sales to be the same of January 2018 to prevent missing observations. By using the `LAG` functions, the query was able to go back a record from the current month, and find the appropriate value of the prior month. If there are no record, a zero will be placed and all results will be formatted as a YYYY-MM date format. Based on the query's result, the month had the largest percentage increase in sales over the prior month is February 2018, with a change in sales of 4.39 times larger than January 2018 (or 439% increase.)

Question 8. Create a VIEW in your own schema that joins together all of the columns in all of the tables. Be aware of Cartesian products e.g., between the ship address and customer.

- SQL:

```
CREATE VIEW Camping_DB
AS
SELECT *
FROM    camping.product p NATURAL JOIN
        camping.orderdetail od NATURAL JOIN
        camping.customerorder o NATURAL JOIN
        camping.rep r NATURAL JOIN
        camping.customersite s NATURAL JOIN
        camping.customer c NATURAL JOIN
        camping.country ct NATURAL JOIN
        camping.branch b;
```

- Result (sample from a 1128 x 49 dimensional dataset):

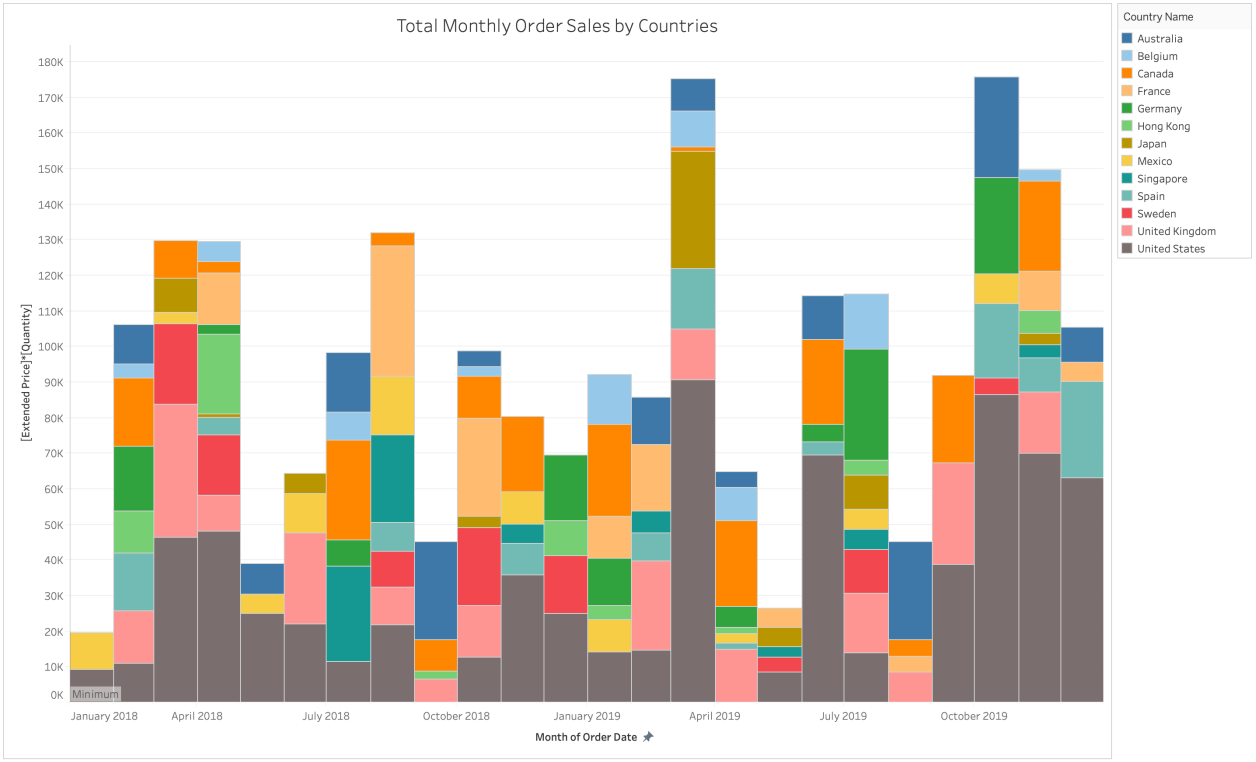
	BRANCH_CD	COUNTRY_CD	CUST_NO	SITE_NO	REP_NO	ORDER_NO	PROD_NO	PROD_TYPE	PROD_LINE	PRODUCT_NAME
1	STOC	SWE	1085	1	1	1001	60400	Environmental Line	Sunblock	Sun Shelter-8
2	STOC	SWE	1085	1	1	1001	60401	Environmental Line	Sunblock	Sun Shelter-15
3	MEX	MEX	1148	1	27	1002	40302	Outdoor Products	Back Packs	GO Small Waist Pack
4	MEX	MEX	1148	1	27	1002	40100	Outdoor Products	Tents	Star Lite
5	MEX	MEX	1148	1	27	1002	40101	Outdoor Products	Tents	Star Gazer-2
6	MEX	MEX	1148	1	27	1002	50101	GO Sport Line	Carry-Bags	GO Ski Gear Bag
7	MEX	MEX	1148	1	27	1002	60400	Environmental Line	Sunblock	Sun Shelter-8
8	MEX	MEX	1148	1	27	1002	60101	Environmental Line	Alert Devices	Microwave Detective
9	MEX	MEX	1030	1	27	1003	40102	Outdoor Products	Tents	Star Gazer-3
10	MEX	MEX	1030	1	27	1003	40100	Outdoor Products	Tents	Star Lite
11	MEX	MEX	1030	1	27	1003	40202	Outdoor Products	Sleeping Bags	MoonLite
12	MEX	MEX	1030	1	27	1003	40103	Outdoor Products	Tents	StarDome
13	MEX	MEX	1030	1	27	1003	50203	GO Sport Line	Sport Wear	GO Water Bottle
14	MEX	MEX	1030	1	27	1003	60301	Environmental Line	Bio-Friendly Soaps	RiverKind Soap
15	MEX	MEX	1030	1	27	1003	60400	Environmental Line	Sunblock	Sun Shelter-8
16	DAL	USA	1038	1	16	1004	50101	GO Sport Line	Carry-Bags	GO Ski Gear Bag
17	DAL	USA	1038	1	16	1004	40103	Outdoor Products	Tents	StarDome
18	DAL	USA	1038	1	16	1004	60500	Environmental Line	Water Purifiers	Pro-Lite Water Filte
19	DAL	USA	1038	1	16	1004	60501	Environmental Line	Water Purifiers	Pocket Water Filter
20	DAL	USA	1038	1	16	1004	40101	Outdoor Products	Tents	Star Gazer-2
21	DAL	USA	1038	1	16	1004	40202	Outdoor Products	Sleeping Bags	MoonLite
22	DAL	USA	1038	1	16	1004	50101	GO Sport Line	Carry-Bags	GO Ski Gear Bag
23	DAL	USA	1038	1	16	1004	60100	Environmental Line	Alert Devices	Pocket U.V. Alerter
24	DAL	USA	1038	1	16	1004	60202	Environmental Line	Recycled Products	Enviro-T
25	TOR	CAN	1093	1	39	1005	40102	Outdoor Products	Tents	Star Gazer-3
26	TOR	CAN	1093	1	39	1005	50202	GO Sport Line	Sport Wear	GO Wristband
27	TOR	CAN	1093	1	39	1005	60100	Environmental Line	Alert Devices	Pocket U.V. Alerter
28	TOR	CAN	1093	1	39	1005	60400	Environmental Line	Sunblock	Sun Shelter-8
29	FRNK	GER	1119	1	9	1006	40403	Outdoor Products	Cooking Equipment	GO Camp Kettle
30	FRNK	GER	1119	1	9	1006	40200	Outdoor Products	Sleeping Bags	MoonBeam
31	FRNK	GER	1119	1	9	1006	40101	Outdoor Products	Tents	Star Gazer-2
32	FRNK	GER	1119	1	9	1006	50101	GO Sport Line	Carry-Bags	GO Ski Gear Bag
33	FRNK	GER	1119	1	9	1006	60402	Environmental Line	Sunblock	Sun Shelter-30
34	FRNK	GER	1119	1	9	1006	60102	Environmental Line	Alert Devices	Pocket Radon Alerter

Question 9. Import all of the VIEW data into Microsoft Excel. Create a pivot table from the resulting dataset and then summarize the data according to total sales by product line and customer state/province.

Total Sales by Provinces/States and Product Lines

Total Sales	Column Labels											
Row Labels	Alert Devices	Back Packs	Bio-Friendly Soaps	Carry-Bags	Cooking Equipment	Recycled Products	Sleeping Bags	Sport Wear	Sunblock	Tents	Water Purifiers	Grand Total
Berkshire	$0	$224	$1,806	$512	$4,745	$0	$0	$0	$1,890	$1,036	$0	$463,443
British Columbia	$620	$361	$24,921	$11,904	$5,143	$312	$1,140	$432	$1,628	$65,130	$311	$3,476,624
California	$189,318	$41,640	$369,603	$196,944	$113,262	$56,110	$76,125	$23,680	$182,825	$1,619,620	$51,675	$83,045,014
Colorado	$40,095	$1,395	$4,114	$560	$8,840	$5,143	$1,920	$6,528	$22,716	$74,800	$18,460	$10,841,584
Florida	$12,710	$96	$26,603	$17,472	$4,320	$792	$1,600	$576	$1,056	$21,437	$2,856	$3,521,807
Illinois	$1,001	$0	$12,948	$1,984	$476	$528	$861	$512	$4,246	$15,939	$0	$965,133
Indiana	$0	$54	$0	$0	$7,326	$66	$0	$128	$1,692	$2,220	$0	$203,616
Manitoba	$1,391	$0	$650	$1,984	$5,100	$168	$5,072	$256	$0	$0	$32,477	$451,557
Maryland	$530	$5,452	$2,145	$560	$0	$2,244	$456	$1,824	$15,180	$3,162	$68,420	$2,508,480
Massachusetts	$3,542	$304	$26,320	$1,488	$8,062	$3,009	$246	$2,448	$0	$4,986	$35,880	$1,915,520
Michigan	$770	$648	$2,873	$2,240	$10,730	$143	$1,677	$1,728	$36,386	$58,240	$16,500	$3,704,175
Missouri	$2,800	$1,014	$46,830	$0	$3,306	$720	$738	$4,752	$0	$500	$311	$1,852,544
New Jersey	$0	$828	$0	$2,800	$1,026	$49	$0	$0	$1,056	$0	$311	$116,272
New York	$4,601	$1,932	$0	$2,976	$8,700	$168	$240	$3,200	$8,679	$4,825	$23,240	$1,067,963
North Carolina	$2,279	$0	$0	$0	$0	$420	$0	$1,088	$9,240	$137,250	$5,320	$1,003,236
Nova Scotia	$1,044	$725	$5,486	$37,632	$1,425	$444	$0	$288	$0	$12,199	$7,455	$2,013,750
NSW	$4,970	$3,969	$2,904	$1,536	$5,100	$12,376	$2,064	$5,184	$1,881	$4,436	$8,370	$2,397,408
Ontario	$56,712	$13,662	$62,181	$9,024	$80,712	$3,717	$29,975	$29,120	$51,964	$29,128	$33,300	$18,766,944
QLD	$20,895	$1,092	$17,472	$11,616	$24,864	$11,118	$17,784	$4,032	$77,116	$17,357	$2,280	$8,554,896
Quebec	$432	$928	$1,749	$24,640	$37,596	$29,808	$24,013	$4,800	$46,480	$51,106	$19,320	$8,542,380
Texas	$10,736	$42	$12,276	$99,060	$12,980	$10,272	$24,752	$25,116	$9,980	$75,582	$184,556	$11,281,456
VIC	$0	$396	$1,518	$4,960	$11,055	$0	$0	$512	$11,778	$2,475	$12,420	$664,353
WA	$5,232	$168	$550	$6,432	$21,021	$0	$2,280	$0	$0	$0	$29,058	$443,864
Wales	$0	$420	$1,116	$4,480	$0	$1,860	$5,727	$0	$0	$0	$0	$114,202
Washington	$0	$1,932	$9,776	$4,224	$3,348	$5,394	$0	$1,008	$17,365	$199,764	$165	$6,565,248
West Sussex	$0	$0	$0	$560	$0	$2,070	$114	$0	$0	$13,048	$190	$77,441
(blank)	$2,706,574	$925,413	$6,393,248	$7,919,881	$5,366,160	$1,054,196	$2,158,830	$840,448	$6,320,314	$28,333,686	$7,069,572	$2,343,667,968
Grand Total	**$12,660,626**	**$4,255,236**	**$25,859,148**	**$27,395,984**	**$22,756,761**	**$5,849,663**	**$8,507,520**	**$4,743,676**	**$25,458,350**	**$108,279,365**	**$32,335,125**	**$9,568,805,184**

Question 10. Import all of the VIEW data into Tableau. Create an appropriate visualization from the resulting data set that summarizes the data according to total sales by product type and customer country over time (e.g., monthly).



Personal Notes.

I carried out the work on my own Macbook Pro; except for the first question, which was done in Baruch computer lab due to the conflict versions between Java Runtime and SQL Developer. I spent around 6 hours working on the assignment. Overall, I found the "justify your rationale" questions are the most difficult since there are no fixed, wrong or correct answers; which makes me feel uncertain about my results.