SQL Case Study

The new ZoomZoom Bat Scooter is now available for sale exclusively through its website. Sales are looking good, but suddenly, pre-orders start plunging by 20% after a couple of weeks. We will figure this out using SQL skills.

We will decompose the problem into the following key steps:

1. Define the question to answer what caused the drop -in sales of the product.
2. Complete background research to gather information to propose an initial hypothesis.
3. Construct a hypothesis to explain the event.
4. Define and execute an objective experiment to test the hypothesis.
5. Analyze the data collected.
6. Report the result of the analysis, which will explain why there was a drop in the sale of Bat Scooters.

Problem Definition: The question that we need to answer is this:

Why did the sales of the ZoomZoom Bat Scoter drop by approximately 20% after about 2 weeks?

Data Collection:

We will collect preliminary data using SQL techniques.

1. We load the sqlda database into PostgreSQL from this link.
2. List the model, base_msrp(Manufacturer's suggested retail price) and production_start_date fields within the product table for product types matching scooter.

```
SELECT model, base_msrp, production_start_date FROM products WHERE prod
uct_type='scooter';
```

Output: Table with details of all products for the scooter product type

Data Output | Explain | Messages | Notifications

	model text	base_msrp numeric	production_start_date timestamp without time zone
1	Lemon	399.99	2010-03-03 00:00:00
2	Lemon Limited Edition	799.99	2011-01-03 00:00:00
3	Lemon	499.99	2013-05-01 00:00:00
4	Blade	699.99	2014-06-23 00:00:00
5	Bat	599.99	2016-10-10 00:00:00
6	Bat Limited Edition	699.99	2017-02-15 00:00:00
7	Lemon Zester	349.99	2019-02-04 00:00:00

3. Looking at the results, we can see that we have two scooter products with Bat in the name: Bat and Bat Limited Edition. Bat Scooter is unique from a price perspective and also from its date of production start. It is the only product with production_start_date in the last quarter. In order to use the sales information, we need to extract the product ID for each of the scooters.

```
SELECT model, product_id FROM products WHERE product_type='scooter';
```

Output:

Data Output | Explain | Messages | Notifications

	model text	product_id bigint
1	Lemon	1
2	Lemon Limited Edition	2
3	Lemon	3
4	Blade	5
5	Bat	7
6	Bat Limited Edition	8
7	Lemon Zester	12

4. Insert the results into a new table called product_names:

```
SELECT model, product_id INTO product_names FROM products WHERE product
_type='scooter';
Select * from product_names;
```

Output:

Data Output | Explain | Messages | Notifications

	model text	product_id bigint
1	Lemon	1
2	Lemon Limited Edition	2
3	Lemon	3
4	Blade	5
5	Bat	7
6	Bat Limited Edition	8
7	Lemon Zester	12

5. Now we will extract sales information for these scooters and examine the sales data.

   Using an inner join on the product_id columns of the tables product_names and sales, we create a new tale called products_sales

```
SELECT model, customer_id, sales_transaction_date, sales_amount, channe
l, dealership_id INTO products_sales FROM sales INNER JOIN product_name
s ON sales.product_id=product_names.product_id;
SELECT * from products_sales LIMIT 5;
```

   Output:

Data Output | Explain | Messages | Notifications

	model text	customer_id bigint	sales_transaction_date timestamp without time zone	sales_amount double precision	channel text	dealership_id double precision
1	Lemon	41604	2012-03-30 22:45:29	399.99	internet	[null]
2	Lemon	41531	2010-09-07 22:53:16	399.99	internet	[null]
3	Lemon	41443	2011-05-24 02:19:11	399.99	internet	[null]
4	Lemon	41291	2010-08-08 14:12:52	319.992	internet	[null]
5	Lemon	41084	2012-01-09 03:34:52	319.992	internet	[null]

6. To look at the first few days of the sales records in detail, we select all the information from the products_sales table for the Bat Scooter and order it by sales_transacion_date in ascending order.

```
SELECT * FROM products_sales WHERE model='Bat' ORDER BY sales_transacti
on_date;
```

   Output:

Data Output | Explain | Messages | Notifications

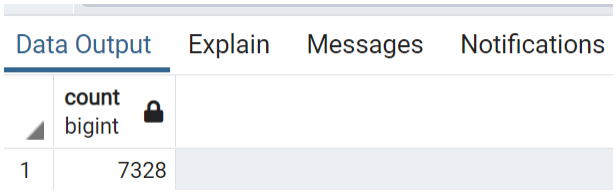
	model text	customer_id bigint	sales_transaction_date timestamp without time zone	sales_amount double precision	channel text	dealership_id double precision
1	Bat	4319	2016-10-10 00:41:57	599.99	internet	[null]
2	Bat	40250	2016-10-10 02:47:28	599.99	dealership	4
3	Bat	35497	2016-10-10 04:21:08	599.99	dealership	2
4	Bat	4553	2016-10-10 07:42:59	599.99	dealership	11
5	Bat	11678	2016-10-10 09:21:08	599.99	internet	[null]
6	Bat	45868	2016-10-10 10:29:29	599.99	internet	[null]
7	Bat	24125	2016-10-10 18:57:25	599.99	dealership	1
8	Bat	31307	2016-10-10 21:22:38	599.99	internet	[null]
9	Bat	42213	2016-10-10 21:27:36	599.99	internet	[null]
10	Bat	47790	2016-10-11 01:28:58	599.99	dealership	20
11	Bat	6342	2016-10-11 03:04:57	599.99	internet	[null]
12	Bat	45880	2016-10-11 04:09:19	599.99	dealership	[null]
13	Bat	43477	2016-10-11 05:24:50	599.99	internet	[null]

✔ Successfully run. Total query runtime: 168 msec. 7328 rows affected.

Count the number of records for the Bat model

```
SELECT COUNT(model) FROM products_sales WHERE model='Bat';
```

Output:



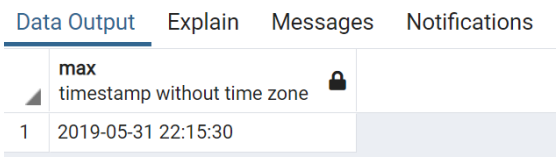
Data Output | Explain | Messages | Notifications

	count bigint
1	7328

7. Determine the last sale date for the Bat Scooter:

```
SELECT MAX(sales_transaction_date) FROM products_sales WHERE model='Bat
';
```

Output:



Data Output | Explain | Messages | Notifications

	max timestamp without time zone
1	2019-05-31 22:15:30

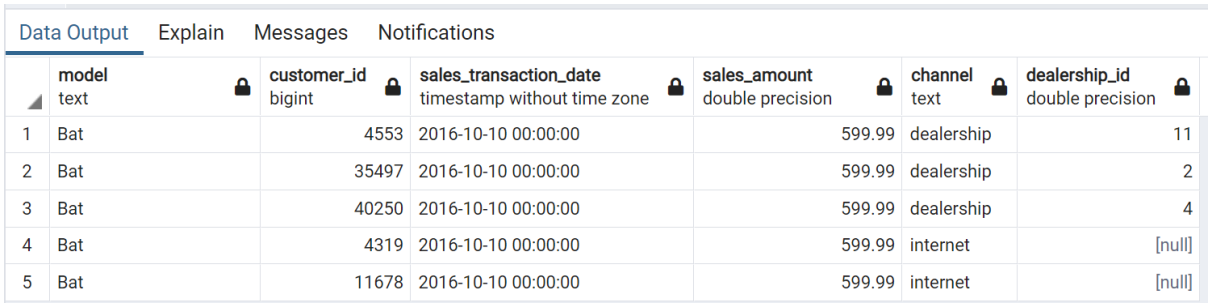
8. Collect the daily sales volume for the Bat Scooter and place in a new table bat_sales to confirm that sales dropped by 20% after the first 2 weeks

```
SELECT * INTO bat_sales FROM products_sales WHERE model='Bat' ORDER BY
sales_transaction_date;
```

9. Remove the time information to allow tracking of sales by date and display first five records of bat_sales

```
UPDATE bat_sales SET sales_transaction_date=DATE(sales_transaction_date
);
SELECT * FROM bat_sales ORDER BY sales_transaction_date LIMIT 5;
```

Output:



Data Output | Explain | Messages | Notifications

	model text	customer_id bigint	sales_transaction_date timestamp without time zone	sales_amount double precision	channel text	dealership_id double precision
1	Bat	4553	2016-10-10 00:00:00	599.99	dealership	11
2	Bat	35497	2016-10-10 00:00:00	599.99	dealership	2
3	Bat	40250	2016-10-10 00:00:00	599.99	dealership	4
4	Bat	4319	2016-10-10 00:00:00	599.99	internet	[null]
5	Bat	11678	2016-10-10 00:00:00	599.99	internet	[null]

10. Create a new table bat_sales_daily containing the sales transaction dates and daily count of total sales and the examine first 22 records (a little over 3 weeks) as sales were reported to have dropped after approximately first 2 weeks

```
SELECT sales_transaction_date, COUNT(sales_transaction_date) INTO batSa
lesDaily FROM bat_sales GROUP BY sales_transaction_date ORDER BY sales_
transaction_date;
SELECT * FROM batSalesDaily LIMIT 22;
```

Output:

Data Output | Explain | Messages | Notifications

	sales_transaction_date timestamp without time zone	count bigint
1	2016-10-10 00:00:00	9
2	2016-10-11 00:00:00	6
3	2016-10-12 00:00:00	10
4	2016-10-13 00:00:00	10
5	2016-10-14 00:00:00	5
6	2016-10-15 00:00:00	10
7	2016-10-16 00:00:00	14
8	2016-10-17 00:00:00	9
9	2016-10-18 00:00:00	11
10	2016-10-19 00:00:00	12
11	2016-10-20 00:00:00	10
12	2016-10-21 00:00:00	6
13	2016-10-22 00:00:00	2
14	2016-10-23 00:00:00	5
15	2016-10-24 00:00:00	6
16	2016-10-25 00:00:00	9
17	2016-10-26 00:00:00	2
18	2016-10-27 00:00:00	4
19	2016-10-28 00:00:00	7
20	2016-10-29 00:00:00	5
21	2016-10-30 00:00:00	5
22	2016-10-31 00:00:00	3

We can see a drop in sales after October 20 as there are 7 days in first 11 rows that record double digit sales and none over next 11 days.

Analysis:

11. Compute the daily cumulative sum of sales and insert into a new table called bat_sales_growth

```
SELECT *, sum(count) OVER (ORDER BY sales_transaction_date) INTO bat_sa
les_growth FROM batSalesDaily;
Select * from bat_sales_growth;
```

Output:

Data Output | Explain | Messages | Notifications

	sales_transaction_date timestamp without time zone	count bigint	sum numeric
1	2016-10-10 00:00:00	9	9
2	2016-10-11 00:00:00	6	15
3	2016-10-12 00:00:00	10	25
4	2016-10-13 00:00:00	10	35
5	2016-10-14 00:00:00	5	40
6	2016-10-15 00:00:00	10	50
7	2016-10-16 00:00:00	14	64
8	2016-10-17 00:00:00	9	73
9	2016-10-18 00:00:00	11	84
10	2016-10-19 00:00:00	12	96
11	2016-10-20 00:00:00	10	106
12	2016-10-21 00:00:00	6	112
13	2016-10-22 00:00:00	2	114
14	2016-10-23 00:00:00	5	119
15	2016-10-24 00:00:00	6	125
16	2016-10-25 00:00:00	9	134
17	2016-10-26 00:00:00	2	136
18	2016-10-27 00:00:00	4	140
19	2016-10-28 00:00:00	7	147
20	2016-10-29 00:00:00	5	152
21	2016-10-30 00:00:00	5	157
22	2016-10-31 00:00:00	3	160

12. Compute a 7-day lag function of sum column and insert all columns of bat_sales_growth and new lag column into a new table to see what sales wer like 1 week before the given record

```
SELECT *, lag(sum, 7) OVER (ORDER BY sales_transaction_date) INTO bat_s
ales_daily_delay FROM bat_sales_growth;
SELECT * FROM bat_sales_daily_delay LIMIT 15;
```

Output:

Data Output | Explain | Messages | Notifications

	sales_transaction_date timestamp without time zone	count bigint	sum numeric	lag numeric
1	2016-10-10 00:00:00	9	9	[null]
2	2016-10-11 00:00:00	6	15	[null]
3	2016-10-12 00:00:00	10	25	[null]
4	2016-10-13 00:00:00	10	35	[null]
5	2016-10-14 00:00:00	5	40	[null]
6	2016-10-15 00:00:00	10	50	[null]
7	2016-10-16 00:00:00	14	64	[null]
8	2016-10-17 00:00:00	9	73	9
9	2016-10-18 00:00:00	11	84	15
10	2016-10-19 00:00:00	12	96	25
11	2016-10-20 00:00:00	10	106	35
12	2016-10-21 00:00:00	6	112	40
13	2016-10-22 00:00:00	2	114	50
14	2016-10-23 00:00:00	5	119	64
15	2016-10-24 00:00:00	6	125	73

13. Compute the sales growth as a percentage and insert the results into new table called bat_sales_delay_vol

```
SELECT *, (sum-lag)/lag AS volume INTO bat_sales_delay_vol FROM bat_sal
es_daily_delay ;
SELECT * FROM bat_sales_daily_delay_vol LIMIT 22;
```

Output:

Data Output | Explain | Messages | Notifications

	sales_transaction_date timestamp without time zone	count bigint	sum numeric	lag numeric	volume numeric
1	2016-10-10 00:00:00	9	9	[null]	[null]
2	2016-10-11 00:00:00	6	15	[null]	[null]
3	2016-10-12 00:00:00	10	25	[null]	[null]
4	2016-10-13 00:00:00	10	35	[null]	[null]
5	2016-10-14 00:00:00	5	40	[null]	[null]
6	2016-10-15 00:00:00	10	50	[null]	[null]
7	2016-10-16 00:00:00	14	64	[null]	[null]
8	2016-10-17 00:00:00	9	73	9	7.1111111111111111
9	2016-10-18 00:00:00	11	84	15	4.6000000000000000
10	2016-10-19 00:00:00	12	96	25	2.8400000000000000
11	2016-10-20 00:00:00	10	106	35	2.0285714285714286
12	2016-10-21 00:00:00	6	112	40	1.8000000000000000
13	2016-10-22 00:00:00	2	114	50	1.2800000000000000
14	2016-10-23 00:00:00	5	119	64	0.85937500000000000000
15	2016-10-24 00:00:00	6	125	73	0.71232876712328767123
16	2016-10-25 00:00:00	9	134	84	0.59523809523809523810
17	2016-10-26 00:00:00	2	136	96	0.41666666666666666667
18	2016-10-27 00:00:00	4	140	106	0.32075471698113207547
19	2016-10-28 00:00:00	7	147	112	0.31250000000000000000
20	2016-10-29 00:00:00	5	152	114	0.33333333333333333333
21	2016-10-30 00:00:00	5	157	119	0.31932773109243697479
22	2016-10-31 00:00:00	3	160	125	0.28000000000000000000

We notice that sales volume on October 17 is 700% above that of launch date of October 10. By October 22, the volume is over double that of the week before. By the end of October the volume is 28% higher than the week prior. So it is confirmed that there is a reduction in sales growth after the first 2 weeks.

14. To try to identify the causes of a sales drop, we will examine the other products in the database.

```
SELECT * FROM products;
```

Output:

Data Output | Explain | Messages | Notifications

	product_id bigint	model text	year bigint	product_type text	base_msrp numeric	production_start_date timestamp without time zone	production_end_date timestamp without time zone
1	1	Lemon	2010	scooter	399.99	2010-03-03 00:00:00	2012-06-08 00:00:00
2	2	Lemon Limited Edition	2011	scooter	799.99	2011-01-03 00:00:00	2011-03-30 00:00:00
3	3	Lemon	2013	scooter	499.99	2013-05-01 00:00:00	2018-12-28 00:00:00
4	5	Blade	2014	scooter	699.99	2014-06-23 00:00:00	2015-01-27 00:00:00
5	7	Bat	2016	scooter	599.99	2016-10-10 00:00:00	[null]
6	8	Bat Limited Edition	2017	scooter	699.99	2017-02-15 00:00:00	[null]
7	12	Lemon Zester	2019	scooter	349.99	2019-02-04 00:00:00	[null]
8	4	Model Chi	2014	automobile	115000.00	2014-06-23 00:00:00	2018-12-28 00:00:00
9	6	Model Sigma	2015	automobile	65500.00	2015-04-15 00:00:00	2018-10-01 00:00:00
10	9	Model Epsilon	2017	automobile	35000.00	2017-02-15 00:00:00	[null]
11	10	Model Gamma	2017	automobile	85750.00	2017-02-15 00:00:00	[null]
12	11	Model Chi	2019	automobile	95000.00	2019-02-04 00:00:00	[null]

We can see that all other products were launched before July except Bat Scooter which was launched in October.

15. List all scooters from the products table

```
SELECT * FROM products WHERE product_type='scooter';
```

Output:

Data Output | Explain | Messages | Notifications

	product_id bigint	model text	year bigint	product_type text	base_msrp numeric	production_start_date timestamp without time zone	production_end_date timestamp without time zone
1	1	Lemon	2010	scooter	399.99	2010-03-03 00:00:00	2012-06-08 00:00:00
2	2	Lemon Limited Edition	2011	scooter	799.99	2011-01-03 00:00:00	2011-03-30 00:00:00
3	3	Lemon	2013	scooter	499.99	2013-05-01 00:00:00	2018-12-28 00:00:00
4	5	Blade	2014	scooter	699.99	2014-06-23 00:00:00	2015-01-27 00:00:00
5	7	Bat	2016	scooter	599.99	2016-10-10 00:00:00	[null]
6	8	Bat Limited Edition	2017	scooter	699.99	2017-02-15 00:00:00	[null]
7	12	Lemon Zester	2019	scooter	349.99	2019-02-04 00:00:00	[null]

Hypothesis Testing:

To test the hypothesis that time of year had an impact on sales, we require a scooter model to use as the control group. Since the Bat Limited Edition Scooter seems to share most of the same features and was launched just 4 months after the Bat scooter, it looks suitable for comparison.

16. Now we carry out similar queries to create a table for Bat Limited Edition scooter to find out its cumulative sales showing volume
    Select first five rows of sales table

```
SELECT * FROM sales LIMIT 5;
```

Output:

Data Output | Explain | Messages | Notifications

	customer_id bigint	product_id bigint	sales_transaction_date timestamp without time zone	sales_amount double precision	channel text	dealership_id double precision
1	1	7	2017-07-19 08:38:41	479.992	internet	[null]
2	22	7	2017-08-14 09:59:02	599.99	dealership	20
3	145	7	2019-01-20 10:40:11	479.992	internet	[null]
4	289	7	2017-05-09 14:20:04	539.991	dealership	7
5	331	7	2019-05-21 20:03:21	539.991	dealership	4

```
SELECT products.model, sales.sales_transaction_date INTO bat_ltd_sales
FROM sales INNER JOIN products ON sales.product_id=products.product_id
WHERE sales.product_id=8 ORDER BY sales.sales_transaction_date;
SELECT * FROM bat_ltd_sales LIMIT 5;
```

Output:

Data Output | Explain | Messages | Notifications

	model text	sales_transaction_date timestamp without time zone
1	Bat Limited Edition	2017-02-15 01:49:02
2	Bat Limited Edition	2017-02-15 09:42:37
3	Bat Limited Edition	2017-02-15 10:48:31
4	Bat Limited Edition	2017-02-15 12:22:41
5	Bat Limited Edition	2017-02-15 13:51:34

```
SELECT COUNT(model) FROM bat_ltd_sales;
```

Data Output

	count bigint
1	5803

```
SELECT MAX(sales_transaction_date) FROM bat_ltd_sales;
```

Data Output | Explain | Messages | Notifications

	max timestamp without time zone
1	2019-05-31 15:08:03

```
ALTER TABLE bat_ltd_sales ALTER COLUMN sales_transaction_date TYPE date
; SELECT * FROM bat_ltd_sales LIMIT 5;
```

Data Output | Explain | Messages | Notifications

	model text	sales_transaction_date date
1	Bat Limited Edition	2017-02-15
2	Bat Limited Edition	2017-02-15
3	Bat Limited Edition	2017-02-15
4	Bat Limited Edition	2017-02-15
5	Bat Limited Edition	2017-02-15

```
SELECT sales_transaction_date, count(sales_transaction_date) INTO bat_l
td_sales_count FROM bat_ltd_sales GROUP BY sales_transaction_date ORDER
BY sales_transaction_date;
SELECT * FROM bat_ltd_sales_count;
```

Data Output | Explain | Messages | Notifications

	sales_transaction_date date	count bigint
1	2017-02-15	6
2	2017-02-16	2
3	2017-02-17	1
4	2017-02-18	4
5	2017-02-19	5
6	2017-02-20	6
7	2017-02-21	5
8	2017-02-22	4
9	2017-02-23	6
10	2017-02-24	2
11	2017-02-25	2
12	2017-02-26	2
13	2017-02-27	4

```
SELECT *, sum(count) OVER (ORDER BY sales_transaction_date) INTO bat_lt
d_sales_growth FROM bat_ltd_sales_count;
SELECT * FROM bat_ltd_sales_growth LIMIT 22;
```

Data Output | Explain | Messages | Notifications

	sales_transaction_date date	count bigint	sum numeric
1	2017-02-15	6	6
2	2017-02-16	2	8
3	2017-02-17	1	9
4	2017-02-18	4	13
5	2017-02-19	5	18
6	2017-02-20	6	24
7	2017-02-21	5	29
8	2017-02-22	4	33
9	2017-02-23	6	39
10	2017-02-24	2	41
11	2017-02-25	2	43
12	2017-02-26	2	45
13	2017-02-27	4	49
14	2017-02-28	4	53
15	2017-03-01	5	58
16	2017-03-02	1	59
17	2017-03-03	3	62
18	2017-03-04	8	70
19	2017-03-05	4	74
20	2017-03-06	7	81
21	2017-03-07	7	88
22	2017-03-08	8	96

Comparing with records of Bat scooter we can see that the Limited Edition scooter did not reach double digits nor did the daily volume of sales fluctuate as much. The Limited edition sold 64 units fewer over first 22 days.

```
SELECT *, lag(sum , 7) OVER (ORDER BY sales_transaction_date) INTO bat_
ltd_sales_delay FROM bat_ltd_sales_growth;
SELECT *, (sum-lag)/lag AS volume INTO bat_ltd_sales_vol FROM bat_ltd_s
ales_delay;
SELECT * FROM bat_ltd_sales_vol LIMIT 22;
```

Data Output | Explain | Messages | Notifications

	sales_transaction_date date	count bigint	sum numeric	lag numeric	volume numeric
1	2017-02-15	6	6	[null]	[null]
2	2017-02-16	2	8	[null]	[null]
3	2017-02-17	1	9	[null]	[null]
4	2017-02-18	4	13	[null]	[null]
5	2017-02-19	5	18	[null]	[null]
6	2017-02-20	6	24	[null]	[null]
7	2017-02-21	5	29	[null]	[null]
8	2017-02-22	4	33	6	4.5000000000000000
9	2017-02-23	6	39	8	3.8750000000000000
10	2017-02-24	2	41	9	3.5555555555555556
11	2017-02-25	2	43	13	2.3076923076923077
12	2017-02-26	2	45	18	1.5000000000000000
13	2017-02-27	4	49	24	1.0416666666666667
14	2017-02-28	4	53	29	0.82758620689655172414
15	2017-03-01	5	58	33	0.75757575757575757576
16	2017-03-02	1	59	39	0.51282051282051282051
17	2017-03-03	3	62	41	0.51219512195121951220
18	2017-03-04	8	70	43	0.62790697674418604651
19	2017-03-05	4	74	45	0.64444444444444444444
20	2017-03-06	7	81	49	0.65306122448979591837
21	2017-03-07	7	88	53	0.66037735849056603774
22	2017-03-08	8	96	58	0.65517241379310344828

Looking at the volume column we can see that the sales growth is more consistent than the Bat scooter. Growth in first week is less but is sutained over a longer period. After 22 days sales growth is 65% compared to previous wek as compared to 28% growth for Bat scooter.

While there is a difference in sales betweenthe two Bat scoters, these can be attributed to the difference in sales price of the two, with Limited edition being $100 moe expensive.

Now we are going to investigate the hypothesis that the reduction in sales gowth could be due to the price point of the Bat scooter. So now we will use the 2013 Lemon scooter for comparison.

17. We carry out similar queries to create a table for 2013 Lemon scooter to find out its cumulative sales showing volume

```
SELECT sales_transaction_date INTO lemon_sales FROM sales WHERE product
_id=3;
SELECT count(sales_transaction_date) FROM lemon_sales;
```

Data Output

	count bigint
1	16558

```
SELECT max(sales_transaction_date) FROM lemon_sales;
```

Data Output | Explain | Messages | Notifications

	max timestamp without time zone
1	2018-12-27 19:12:10

```
ALTER TABLE lemon_sales ALTER COLUMN sales_transaction_date TYPE DATE;
SELECT *, COUNT(sales_transaction_date) INTO lemon_sales_count FROM lem
on_sales GROUP BY sales_transaction_date ORDER BY sales_transaction_dat
e;
SELECT *, sum(count) OVER (ORDER BY sales_transaction_date) INTO lemon_
sales_sum FROM lemon_sales_count;
SELECT *, lag(sum, 7) OVER (ORDER BY sales_transaction_date) INTO lemon
_sales_delay FROM lemon_sales_sum;
SELECT *, (sum-lag)/lag AS volume INTO lemon_sales_growth FROM lemon_sa
les_delay;
SELECT * FROM lemon_sales_growth LIMIT 22;
```

Data Output | Explain | Messages | Notifications

	sales_transaction_date date	count bigint	sum numeric	lag numeric	volume numeric
1	2013-05-01	6	6	[null]	[null]
2	2013-05-02	8	14	[null]	[null]
3	2013-05-03	4	18	[null]	[null]
4	2013-05-04	9	27	[null]	[null]
5	2013-05-05	9	36	[null]	[null]
6	2013-05-06	6	42	[null]	[null]
7	2013-05-07	8	50	[null]	[null]
8	2013-05-08	6	56	6	8.3333333333333333
9	2013-05-09	6	62	14	3.4285714285714286
10	2013-05-10	9	71	18	2.9444444444444444
11	2013-05-11	3	74	27	1.7407407407407407
12	2013-05-12	4	78	36	1.1666666666666667
13	2013-05-13	7	85	42	1.0238095238095238
14	2013-05-14	3	88	50	0.76000000000000000000
15	2013-05-15	3	91	56	0.62500000000000000000
16	2013-05-16	4	95	62	0.53225806451612903226
17	2013-05-17	6	101	71	0.42253521126760563380
18	2013-05-18	9	110	74	0.48648648648648648649
19	2013-05-19	6	116	78	0.48717948717948717949
20	2013-05-20	6	122	85	0.43529411764705882353
21	2013-05-21	11	133	88	0.51136363636363636364
22	2013-05-22	8	141	91	0.54945054945054945055

We can see that the initial sales volume is much higher than the other scooters at over 80% and finishes higher at 55%.

			Sales growth	
Product	**Launched**	**Price**	**Second week**	**Day 22**
2013 Lemon	May-13	$499.99	830%	55%
Bat original	Oct-17	$599.99	700%	28%
Bat Limited edition	Feb-18	$699.99	450%	66%

Now we have extracted data to test the two hypothesis of timing and cost. Based on the information collected we can make the folowing conclusions:

1. The initial growth rate starting in the second week of sales negatively correlates to the cost of the scooter. As the cost increased the initial growth rate dropped.
2. There is some evidence to suggest that the reduction in sales could be due to seasonal variations given the significant reduction in growth and the fact that the original Bat scooter was the only one released in October. So the drop can be attributed to the difference in launch timing.

Now we will analyze the opening rate of marketing emails and find out if it has an effect on sales growth throughout the first 3 weeks.

First, look at the emails table to see what information is available.

```
SELECT * FROM emails LIMIT 5;
```

Data Output | Explain | Messages | Notifications

	email_id bigint	customer_id bigint	email_subject text	opened text	clicked text	bounced text	sent_date timestamp without time zone	opened_date timestamp without time zone	clicked_date timestamp without time zone
1	1	18	Introducing A Limited Edition	f	f	f	2011-01-03 15:00:00	[null]	[null]
2	2	30	Introducing A Limited Edition	f	f	f	2011-01-03 15:00:00	[null]	[null]
3	3	41	Introducing A Limited Edition	t	f	f	2011-01-03 15:00:00	2011-01-04 10:41:11	[null]
4	4	52	Introducing A Limited Edition	f	f	f	2011-01-03 15:00:00	[null]	[null]
5	5	59	Introducing A Limited Edition	f	f	f	2011-01-03 15:00:00	[null]	[null]

We need to know whether an email was opened, and when it was opened , who the customer was and whether they purchased a scooter.

1. As we want the email records of customers who purchased a Bat scooter, we will join the customer_id column in tables emails and bat_sales and insert the results into a new table bat_emails.

```
SELECT emails.email_subject, emails.customer_id, emails.opened, emails.
sent_date, emails.opened_date, bat_sales.sales_transaction_date INTO ba
t_emails FROM emails INNER JOIN bat_sales ON bat_sales.customer_id=emai
ls.customer_id ORDER BY bat_sales.sales_transaction_date;
SELECT * FROM bat_emails LIMIT 10;
```

Data Output | Explain | Messages | Notifications

	email_subject text	customer_id bigint	opened text	sent_date timestamp without time zone	opened_date timestamp without time zone	sales_transaction_date timestamp without time zone
1	Save the Planet with some Ho...	31307	f	2018-11-23 15:00:00	[null]	2016-10-10 00:00:00
2	Zoom Zoom Black Friday Sale	31307	f	2014-11-28 15:00:00	[null]	2016-10-10 00:00:00
3	Like a Bat out of Heaven	24125	f	2016-09-21 15:00:00	[null]	2016-10-10 00:00:00
4	Like a Bat out of Heaven	11678	f	2016-09-21 15:00:00	[null]	2016-10-10 00:00:00
5	Like a Bat out of Heaven	42213	f	2016-09-21 15:00:00	[null]	2016-10-10 00:00:00
6	A Brand New Scooter...and Car	24125	f	2014-05-06 15:00:00	[null]	2016-10-10 00:00:00
7	Like a Bat out of Heaven	40250	f	2016-09-21 15:00:00	[null]	2016-10-10 00:00:00
8	An Electric Car for a New Age	31307	f	2015-04-01 15:00:00	[null]	2016-10-10 00:00:00
9	A New Year, And Some New E...	4553	f	2019-01-07 15:00:00	[null]	2016-10-10 00:00:00
10	Shocking Holiday Savings On ...	40250	f	2013-11-29 15:00:00	[null]	2016-10-10 00:00:00

2. Select all rows where sent_date is before the sales_transaction_date.

```
SELECT * FROM bat_emails WHERE sent_date < sales_transaction_date ORDER
BY customer_id LIMIT 22;
```

Data Output | Explain | Messages | Notifications

	email_subject text	customer_id bigint	opened text	sent_date timestamp without time zone	opened_date timestamp without time zone	sales_transaction_date timestamp without time zone
1	The 2013 Lemon Scooter is H...	7	f	2013-03-01 15:00:00	[null]	2019-04-25 00:00:00
2	Zoom Zoom Black Friday Sale	7	f	2014-11-28 15:00:00	[null]	2019-04-25 00:00:00
3	Save the Planet with some Ho...	7	f	2018-11-23 15:00:00	[null]	2019-04-25 00:00:00
4	Like a Bat out of Heaven	7	f	2016-09-21 15:00:00	[null]	2019-04-25 00:00:00
5	Tis' the Season for Savings	7	f	2015-11-26 15:00:00	[null]	2019-04-25 00:00:00
6	We Really Outdid Ourselves th...	7	f	2017-01-15 15:00:00	[null]	2019-04-25 00:00:00
7	25% off all EVs. It's a Christm...	7	t	2016-11-25 15:00:00	2016-11-26 03:55:30	2019-04-25 00:00:00
8	A New Year, And Some New E...	7	f	2019-01-07 15:00:00	[null]	2019-04-25 00:00:00
9	A Brand New Scooter...and Car	7	f	2014-05-06 15:00:00	[null]	2019-04-25 00:00:00
10	Black Friday. Green Cars.	7	f	2017-11-24 15:00:00	[null]	2019-04-25 00:00:00
11	Shocking Holiday Savings On ...	7	f	2013-11-29 15:00:00	[null]	2019-04-25 00:00:00
12	We cut you a deal: 20%% off a...	7	t	2014-09-18 15:00:00	2014-09-19 15:11:17	2019-04-25 00:00:00
13	An Electric Car for a New Age	7	t	2015-04-01 15:00:00	2015-04-02 15:10:55	2019-04-25 00:00:00
14	An Electric Car for a New Age	22	f	2015-04-01 15:00:00	[null]	2017-08-14 00:00:00
15	A Brand New Scooter...and Car	22	t	2014-05-06 15:00:00	2014-05-07 13:31:23	2017-08-14 00:00:00
16	Zoom Zoom Black Friday Sale	22	t	2014-11-28 15:00:00	2014-11-29 11:31:03	2017-08-14 00:00:00
17	The 2013 Lemon Scooter is H...	22	f	2013-03-01 15:00:00	[null]	2017-08-14 00:00:00
18	We cut you a deal: 20%% off a...	22	f	2014-09-18 15:00:00	[null]	2017-08-14 00:00:00
19	25% off all EVs. It's a Christm...	22	f	2016-11-25 15:00:00	[null]	2017-08-14 00:00:00
20	Shocking Holiday Savings On ...	22	f	2013-11-29 15:00:00	[null]	2017-08-14 00:00:00
21	Like a Bat out of Heaven	22	f	2016-09-21 15:00:00	[null]	2017-08-14 00:00:00
22	Tis' the Season for Savings	22	f	2015-11-26 15:00:00	[null]	2017-08-14 00:00:00

3. We can see that there are some rows where emails are sent years before the transaction date. We can remove those by deleting the rows where emails were sent more than 6 months prior to production.

```
DELETE FROM bat_emails WHERE sent_date < '2016-04-10';
```

4. Also delete rows where emails were sent after the purchase

```
DELETE FROM bat_emails WHERE sent_date > sales_transaction_date;
```

5. Delete those rows where email sent was within a month prior to the purchase date. Emails sent much earlier are unlikely to influence a purchase decision.

```
DELETE FROM bat_emails WHERE (sales_transaction_date-sent_date) > '30 d
ays';
SELECT * FROM bat_emails ORDER BY customer_id LIMIT 22;
```

Data Output | Explain | Messages | Notifications

	email_subject text	customer_id bigint	opened text	sent_date timestamp without time zone	opened_date timestamp without time zone	sales_transaction_date timestamp without time zone
1	25% off all EVs. It's a Christm...	129	t	2016-11-25 15:00:00	2016-11-26 06:31:37	2016-11-28 00:00:00
2	A New Year, And Some New E...	145	f	2019-01-07 15:00:00	[null]	2019-01-20 00:00:00
3	Black Friday. Green Cars.	150	f	2017-11-24 15:00:00	[null]	2017-12-19 00:00:00
4	Black Friday. Green Cars.	173	f	2017-11-24 15:00:00	[null]	2017-12-05 00:00:00
5	We Really Outdid Ourselves th...	196	f	2017-01-15 15:00:00	[null]	2017-01-23 00:00:00
6	We Really Outdid Ourselves th...	319	f	2017-01-15 15:00:00	[null]	2017-01-29 00:00:00
7	Like a Bat out of Heaven	369	f	2016-09-21 15:00:00	[null]	2016-10-13 00:00:00
8	Like a Bat out of Heaven	414	f	2016-09-21 15:00:00	[null]	2016-10-20 00:00:00
9	25% off all EVs. It's a Christm...	418	f	2016-11-25 15:00:00	[null]	2016-12-21 00:00:00
10	A New Year, And Some New E...	560	t	2019-01-07 15:00:00	2019-01-08 15:56:14	2019-01-29 00:00:00
11	We Really Outdid Ourselves th...	600	f	2017-01-15 15:00:00	[null]	2017-01-18 00:00:00
12	A New Year, And Some New E...	660	t	2019-01-07 15:00:00	2019-01-08 23:37:03	2019-01-08 00:00:00
13	A New Year, And Some New E...	681	f	2019-01-07 15:00:00	[null]	2019-01-13 00:00:00
14	Black Friday. Green Cars.	806	t	2017-11-24 15:00:00	2017-11-25 16:59:40	2017-11-29 00:00:00
15	A New Year, And Some New E...	881	t	2019-01-07 15:00:00	2019-01-08 21:07:28	2019-01-22 00:00:00
16	25% off all EVs. It's a Christm...	934	t	2016-11-25 15:00:00	2016-11-26 09:22:45	2016-12-24 00:00:00
17	25% off all EVs. It's a Christm...	983	f	2016-11-25 15:00:00	[null]	2016-11-29 00:00:00
18	A New Year, And Some New E...	1060	f	2019-01-07 15:00:00	[null]	2019-01-27 00:00:00
19	25% off all EVs. It's a Christm...	1288	f	2016-11-25 15:00:00	[null]	2016-12-11 00:00:00
20	25% off all EVs. It's a Christm...	1317	f	2016-11-25 15:00:00	[null]	2016-12-13 00:00:00
21	A New Year, And Some New E...	1400	t	2019-01-07 15:00:00	2019-01-08 15:01:00	2019-01-10 00:00:00
22	Save the Planet with some Ho...	1417	f	2018-11-23 15:00:00	[null]	2018-11-26 00:00:00

6. There are a few emails that are unrelated to the Bat scooter. We can delete those.

```
SELECT DISTINCT(email_subject) FROM bat_emails;
```

Data Output | Explain | Messages | Notifications

	email_subject text
1	Black Friday. Green Cars.
2	25% off all EVs. It's a Christm...
3	A New Year, And Some New E...
4	Like a Bat out of Heaven
5	Save the Planet with some Ho...
6	We Really Outdid Ourselves th...

```
DELETE FROM bat_emails WHERE position('Black Friday' in email_subject)>
0;
DELETE FROM bat_emails WHERE position('25% off all EV' in email_subject
)>0;
DELETE FROM bat_emails WHERE position('Some New EV' in email_subject)>0
;
SELECT count(sales_transaction_date) FROM bat_emails;
```

Data Output

	count bigint
1	401

There are 401 rows left in the sample.

7. Find the percentage of emails opened relative to sales.

```
SELECT count(opened) FROM bat_emails WHERE opened='t'
```

Data Output

	count bigint
1	98

8. Count the customers who received emails and made a purchase.

```
SELECT COUNT(DISTINCT(customer_id)) FROM bat_emails;
```

Data Output

	count bigint
1	396

9. Count the unique customers who made a purchase

```
SELECT COUNT(DISTINCT(customer_id)) FROM bat_sales;
```

Data Output

	count bigint
1	6659

```
SELECT 396.0/6659.0 AS email_rate;
```

Data Output | Explain | Messages | Notifications

	email_rate numeric
1	0.05946838864694398558

So we can see that just under 6% customers who made a purchase received an email regarding the Bat Scooter. There is a strong argument that actively increasing theg size of the customer base who receive marketing emails could increase the Bat scooter sales.

If we limit the scope of our data to all sales within first 3 weeks and check the rate of opening of emails where we saw a reduction in sales.

```
SELECT * INTO bat_emails_threewks FROM bat_emails WHERE sales_transacti
on_date < '2016-11-01';
SELECT COUNT(opened) FROM bat_emails_threewks;
```

Data Output

	count bigint
1	82

```
SELECT COUNT(opened) FROM bat_emails_threewks WHERE opened='t';
```

Data Output

	count bigint
1	15

```
SELECT 15.0/82.0 AS sale_rate;
```

Data Output | Explain | Messages | Notifications

	sale_rate numeric
1	0.18292682926829268293

So approximately 18% customers who received an email about the Bat scooter made a purchase in the first 3 weeks.
Find how many unique customers are there in total throughout first 3 weeks.

```
SELECT COUNT(DISTINCT(customer_id)) FROM bat_sales WHERE sales_transact
ion_date < '2016-11-01';
```

Data Output

	count bigint
1	160

There are 160 customers in the first 3 weeks, 82 of whom received emails, which is slightly over 50% of customers. This is much more than 6% of customers over the entire period of availability of the scooter.

Conclusions:

1. In Launch Timing Analysis, we gathered some evidence to suggest that launch timing could be related to reduction in sales after first 2 weeks.
2. There is a correlation between the initial sales rate and the sales price of the scooter, with a reduced-sales price trending with high sales rate.
3. The number of units sold in the first 3 weeks does not directly correlate to the sale price of the product.
4. There is evidence to suggest that a successful marketing campaign could increase the initial sales rate, with an increased email opening rate trending with an increased sales rate.
5. The Bat scooter sold more units in the first 3 weeks than the Lemon or the Bat limited edition scoters.