Data Analysis with SQL

Time

Example: Insert

```
CREATE TABLE people (
      name VARCHAR(20),
      dob  DATE
);

INSERT INTO people VALUES ('Larry David', '1947-07-02');
```

Format

A date/time string can be (inserted/stored) in many formats:

PostgreSQL: Documentation: 13: 9.8. Data Type Formatting Functions

Examples

```
INSERT INTO people VALUES ('Larry David', '1947-07-02');

INSERT INTO people VALUES ('James Thomson', to_date('14 Dec 2002', 'DD Mon YYYY'));

INSERT INTO people VALUES ('Ben Stiller', to_date('November 30, 1965', 'Month DD, YYYY'));


select name, to_char(dob,'Month DD, YYYY')
from people;
```

```
Larry David,          "July      02, 1947"
James Thomson,        "December  14, 2002"
Ben Stiller,          "November  30, 1965"
```

Extract info from date

```
SELECT EXTRACT(YEAR FROM dob)
FROM people;

SELECT EXTRACT(MONTH FROM dob)
FROM people;

--extract day of the month
SELECT EXTRACT(DAY FROM dob)
FROM people;

--extract day of the year
SELECT EXTRACT(DOY FROM dob)
FROM people;

--extract day of week Sunday (0) to Saturday (6)
SELECT EXTRACT(DOW FROM dob)
FROM people;

SELECT EXTRACT(WEEK FROM dob)
FROM people;

SELECT EXTRACT(QUARTER FROM dob)
FROM people;


SELECT name, age(dob)
FROM people;
```

Orders, Orderlines, Products

```
CREATE TABLE orders (
    orderid int,
    customerid int,
    campaignid int,
    orderdate date,
    city varchar(50),
    state char(2),
    zipcode char(5),
    paymenttype varchar(50),
    totalprice int,
    numorderlines int,
    numunits int
);
```

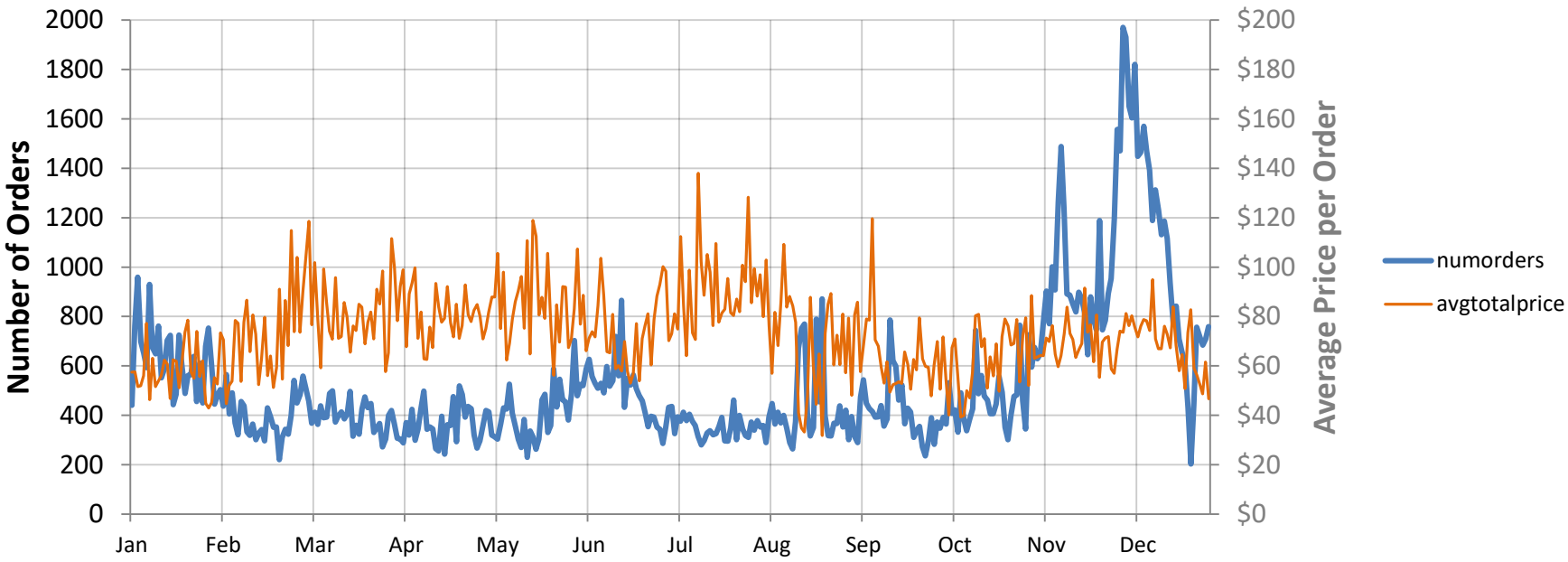
```
CREATE TABLE orderline (
    orderlineid int,
    orderid int,
    productid int,
    shipdate date,
    billdate date,
    unitprice real,
    numunits int,
    totalprice real
);
```

```
CREATE TABLE products (
        productid int,
        productname text,
        productgroupcode text,
        productgroupname
        text,
        instockflag text,
        fullprice int
);
```

How do the number of orders and average order price vary by day of the year?

```
SELECT
        EXTRACT(month from orderdate) AS month,
        EXTRACT(day from orderdate) AS day,
        COUNT(orderid) AS numorders,
        AVG(totalprice) AS avgtotalprice
FROM orders
GROUP BY month, day
ORDER BY 1;
```

Chart


Number of Orders
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
$200
$180
$160
$140
$120
$100
$80
$60
$40
$20
$0
Average Price per Order
numorders
avgtotalprice

How many orders are placed on each day of the week?

```
SELECT
    EXTRACT(dow from orderdate) AS dayofweek,
    COUNT(orderid) AS numorders
FROM orders
GROUP BY dayofweek
ORDER BY 1;
```

How many orders are placed on each day of the week? – **Horizontally**

```
SELECT
    SUM(CASE WHEN EXTRACT(dow from orderdate)=0 THEN 1 ELSE 0 END) AS Sun,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=1 THEN 1 ELSE 0 END) AS Mon,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=2 THEN 1 ELSE 0 END) AS Tue,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=3 THEN 1 ELSE 0 END) AS Wed,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=4 THEN 1 ELSE 0 END) AS Thu,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=5 THEN 1 ELSE 0 END) AS Fri,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=6 THEN 1 ELSE 0 END) AS Sat
FROM orders;
```

Has the number of orders by day of the week changed over the years?

```
SELECT EXTRACT(year from orderdate) AS theyear,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=0 THEN 1 ELSE 0 END) AS Sun,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=1 THEN 1 ELSE 0 END) AS Mon,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=2 THEN 1 ELSE 0 END) AS Tue,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=3 THEN 1 ELSE 0 END) AS Wed,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=4 THEN 1 ELSE 0 END) AS Thu,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=5 THEN 1 ELSE 0 END) AS Fri,
    SUM(CASE WHEN EXTRACT(dow from orderdate)=6 THEN 1 ELSE 0 END) AS Sat
FROM orders
GROUP BY theyear;
```

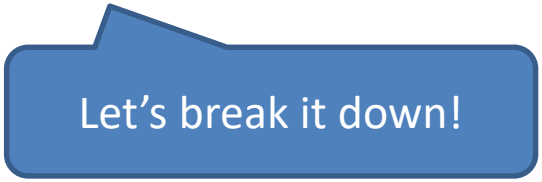
Alternative + Excel Pivot

```
SELECT
    EXTRACT(year from orderdate) AS theyear,
    EXTRACT(dow from orderdate) AS dayofweek,
    COUNT(orderid) AS cnt
FROM orders
GROUP BY theyear, dayofweek;
```

What is the number of orders by quarter of each year?

```
SELECT
      EXTRACT(year from orderdate) AS yr,
      EXTRACT(quarter from orderdate) AS quarter,
  COUNT(orderid) AS numorders
FROM orders
GROUP BY yr, quarter
ORDER BY 1,2;
```

What is the product category of the most popular product during each month?



What is the frequency of each product in each month?

```
SELECT
      EXTRACT(year from orderdate) as yr,
      EXTRACT(month from orderdate) as mon,
      productid,
      COUNT(*) as cnt
FROM orders JOIN orderline USING(orderid)
GROUP BY yr, mon, productid;
```

The date is in **orders**, whereas product is in **orderline**. So, we need to join **orders** with **orderline**.

We will produce the result in vertical form.

Q1

Results

yr	mon	productid	cnt
2009	10	10012	11
2009	10	10013	4
2009	10	10014	1
2009	10	10015	4
2009	10	10109	3
2009	10	10127	1
2009	10	10134	4
2009	10	10382	2
2009	10	10384	1
…	…	…	…

What is the maximum frequency in each month?

```
SELECT yr, mon, MAX(cnt) AS maxcnt
FROM (
    Q1
)
GROUP BY yr, mon;
```

Q2

Results

yr	mon	maxcnt
2009	10	1063
2009	11	2353
2009	12	658
2010	1	226
2010	2	47
2010	3	97
2010	4	180
2010	5	145
2010	6	202
…	…	…

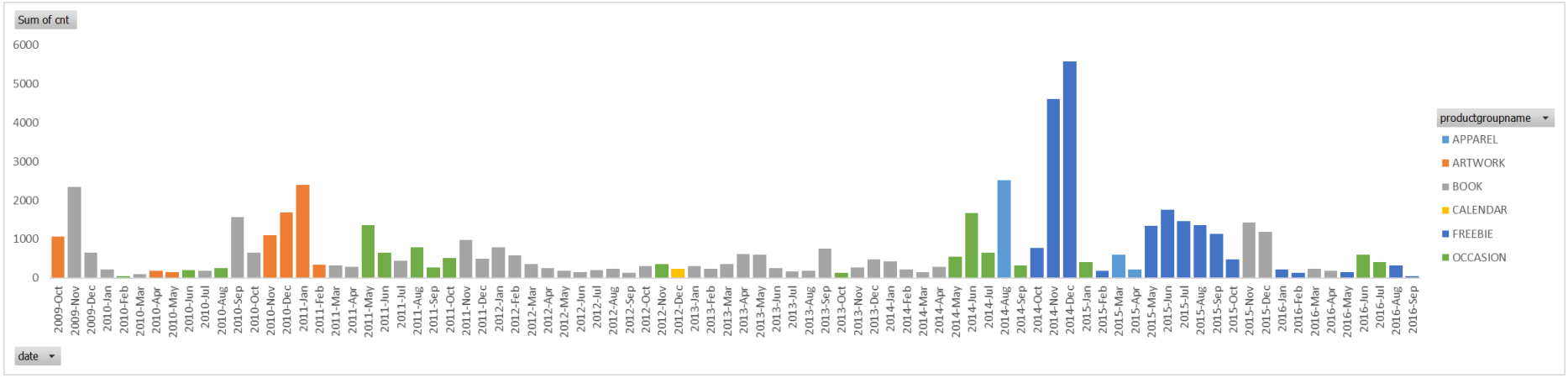
What is the product category of the most popular product during each month?

```
SELECT prodmon.yr, prodmon.mon, prodmon.cnt, p.productgroupname
FROM  Q1 prodmon
        JOIN Q2 prodmax ON
                prodmon.yr = prodmax.yr AND
                prodmon.mon = prodmax.mon AND
                prodmon.cnt = prodmax.maxcnt
        JOIN products p ON
                 prodmon.productid = p.productid
ORDER BY 1, 2;
```

Results

yr	mon	cnt	PRODUCTGROUPNAME
2009	10	1063	ARTWORK
2009	11	2353	BOOK
2009	12	658	BOOK
2010	1	226	BOOK
2010	2	47	OCCASION
2010	3	97	BOOK
2010	4	180	ARTWORK
2010	5	145	ARTWORK
2010	6	202	OCCASION
…	…	…	…

Chart


Sum of cnt
6000
5000
4000
3000
2000
1000
0
2009-Oct
2009-Nov
2009-Dec
2010-Jan
2010-Feb
2010-Mar
2010-Apr
2010-May
2010-Jun
2010-Jul
2010-Aug
2010-Sep
2010-Oct
2010-Nov
2010-Dec
2011-Jan
2011-Feb
2011-Mar
2011-Apr
2011-May
2011-Jun
2011-Jul
2011-Aug
2011-Sep
2011-Oct
2011-Nov
2011-Dec
2012-Jan
2012-Feb
2012-Mar
2012-Apr
2012-May
2012-Jun
2012-Jul
2012-Aug
2012-Sep
2012-Oct
2012-Nov
2012-Dec
2013-Jan
2013-Feb
2013-Mar
2013-Apr
2013-May
2013-Jun
2013-Jul
2013-Aug
2013-Sep
2013-Oct
2013-Nov
2013-Dec
2014-Jan
2014-Feb
2014-Mar
2014-Apr
2014-May
2014-Jun
2014-Jul
2014-Aug
2014-Sep
2014-Oct
2014-Nov
2014-Dec
2015-Jan
2015-Feb
2015-Mar
2015-Apr
2015-May
2015-Jun
2015-Jul
2015-Aug
2015-Sep
2015-Oct
2015-Nov
2015-Dec
2016-Jan
2016-Feb
2016-Mar
2016-Apr
2016-May
2016-Jun
2016-Jul
2016-Aug
2016-Sep
date
productgroupname
APPAREL
ARTWORK
BOOK
CALENDAR
FREEBIE
OCCASION