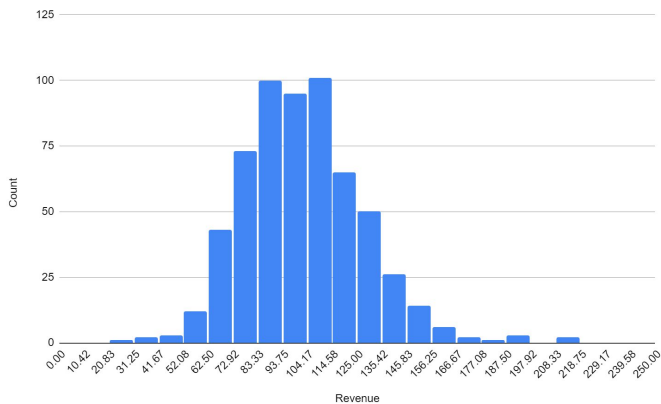


What is the Distribution of Customer Lifetime Value

Histogram of Life Time Value



SQL QUERY

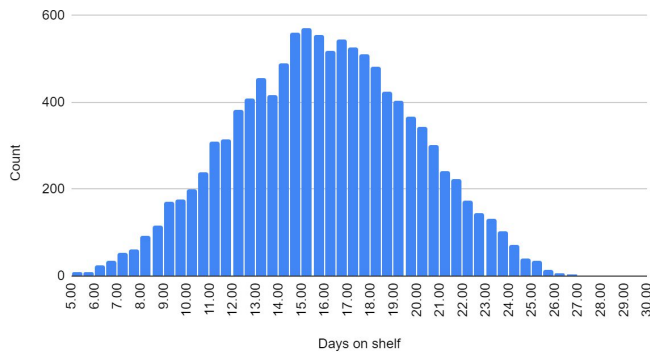
```
with t1 as (SELECT date_trunc('month', p.payment_date) as pay_mon,
                  c.first_name || ' ' || c.last_name as
                  fullname,
                  p.customer_id,
                  p.amount,
                  p.payment_id
FROM payment as p
JOIN customer as c
ON p.customer_id = c.customer_id),
t2 as (SELECT pay_mon, fullname,
              count(payment_id) OVER (partition by
              pay_mon, fullname
              order by fullname) as pay_counterpermon,
              sum(amount) OVER (partition by pay_mon, fullname
              order by fullname) as pay_amount,
              sum(amount) over (partition by fullname order by
              fullname) as running_total
FROM t1
order by 5 desc, 2, 1)

SELECT fullname, running_total
FROM t2
group by 1, 2
```

What we are looking at here is a distribution for customer lifetime value or the running total amount of revenue generated by all customers. Currently we can see the population of customers is grouped rather tightly grouped.

What is the Distribution of time a dvd spends on the shelf

Histogram of Days on the shelf

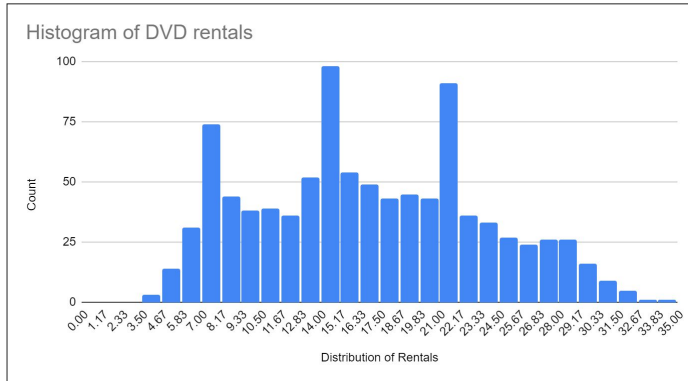


SQL QUERY

```
with t1 as (SELECT f.title, count(*)
            FROM film as f
            JOIN inventory as i
            on i.film_id = f.film_id
            group by f.title),
t2 as (SELECT f.title, i.inventory_id, r.rental_date, r.return_date,
            LEAD(rental_date) over (partition by
            i.inventory_id order by i.inventory_id, rental_date) - return_date as
            time_on_shelf
            FROM film as f
            JOIN inventory as i
            on i.film_id = f.film_id
            JOIN rental as r
            ON i.inventory_id = r.inventory_id
            order by 2,3)
--t3 as (average rental time)
SELECT inventory_id,
       (extract(epoch from time_on_shelf)/(3600*24))::decimal(4,2) as
       days_on_shelf
FROM t2
where extract(epoch from time_on_shelf)/(3600*24) is not null and
       extract(epoch from time_on_shelf)/(3600*24) < 29
```

First off some outliers have been removed as there are a few films that sat on the shelf for over 100 days! What we see here is a pretty even distribution of how long a movie sits on the self. Perhaps we should aimat reducing the time a movie sits on the shelf? Such as selling dvds at deep discount that sit on the shelf for longer than a 20 day period without being rented?

<What is the Distribution of Movie Demand by Title>



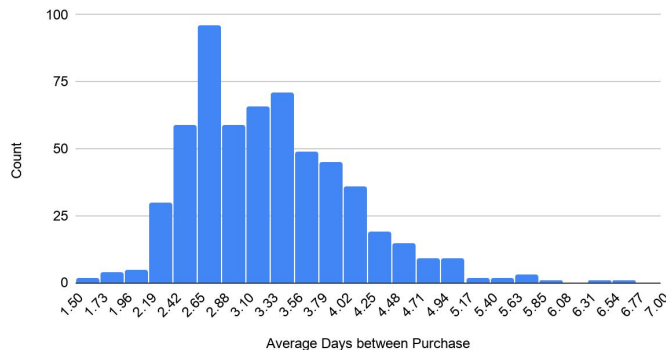
SQL QUERY

```
SELECT f.title, count(r.*)
FROM rental as r
JOIN inventory as i
ON i.inventory_id = r.inventory_id
JOIN film as f
ON i.film_id = f.film_id
GROUP BY f.title
order by 2 desc
```

Here is a simple histogram showing the distribution of movie demand by title. There are ~75 titles that get rented 7 times, ~95 titles that have been rented out 14 times, and ~85 titles that have been rented out a whopping 21 times! Some dvds get rented out a lot!

<What is the Distribution for the Average time between purchases for a customer>

Histogram of Avg days between Purchases



SQL QUERY

```
with t1 as (SELECT c.first_name || ' ' || c.last_name as fullname,
                  payment_date
FROM payment as p
JOIN customer as c
ON p.customer_id = c.customer_id
WHERE c.active = 1),

t2 as (SELECT *, payment_date - LAG(payment_date) OVER
(partition by fullname order by fullname, payment_date) as re_occurring
FROM t1)

SELECT DISTINCT fullname,
extract(epoch from (avg(re_occurring) OVER (partition by
fullname order by fullname))) / (3600 * 24) as avg_day_to_purchase
FROM t2
order by 2
```

This shows the average time between rentals by customer. What we see is that there is a sharp increase in average time between purchases. Very few customers average less than 2 days between purchases, the majority return between 2 and 4 days. Could it be that customers are making purchases when they return 2 day or 4 day rentals?