```
# Load conditions as nodes
load csv with headers from "file:///conditions.csv" as sample
with sample
merge (c: Condition {condition_name: sample.`condition`})
return c

# Load drugs as nodes
load csv with headers from "file:///drug_info.csv" as sample
with sample
merge (d: Drug {drug_name: sample.`drugNameReduced`, effectiveness: sample.`effectiveness`,
sideEffects: sample.`sideEffects`, rating: sample.`rating`, sentiment_score:
sample.`sentiment_score`, benefits_score: sample.`benefits_sentiment_score`,
side_effects_sentiment_score: sample.`side_effects_sentiment_score`})
return d

#Load genes as nodes
load csv with headers from "file:///genes.csv" as sample
with sample
merge (g: Gene{gene_name: sample.`gene`,gene_full_name: sample.`gene_name`, chromosome:
sample.`chromosome`, chromosome_band: sample.`chromosome_band`, dna_start:
sample.`dna_start`, dna_end: sample.`dna_end`})
return g

# Load edges between drugs and the conditions they treat
load csv with headers from "file:///treatment.csv" as row
match (c: Condition {condition_name: row.`condition`})
match (d: Drug {drug_name: row.`drug`})
merge (d)-[:TREATS]->(c)

# Load edges between genes and their associated conditions
load csv with headers from "file:///gene_interactions.csv" as row
match (c: Condition {condition_name: row.`condition`})
match (g: Gene {gene_name: row.`gene`})
merge (g)-[:ASSOCIATED]->(c)

# Return connection between drugs, gene, and conditions
match (g:Gene)--(c:Condition)--(d:Drug)--(x:Condition)--(y:Gene)
return g,c,d,x,y
limit 25
```

For a given condition, are there already numerous drugs on the market that are positively regarded by users?

Initial exploration with acne:

```
# Return top 10 drugs that treat acne based on rating
match (c:Condition {condition_name: 'acne'})-[r:TREATS]-(d:Drug)
return c, d, d.rating as rating
order by rating desc
limit 10
```

```
# Return graph of acne drugs that have a rating greater than 8.5
match (c:Condition {condition_name: 'acne'})-[r:TREATS]-(d:Drug)
where toFloat(d.rating) > 8.5
return d.drug_name, d.rating as rating
order by toFloat(rating) desc
```

Generalize for the market by only looking into conditions that have more than 20 associated drugs:

```
# Return the highest average drug ratings and number of associated drugs for conditions
match (c:Condition)-[r:TREATS]-(d:Drug)
with  c.condition_name as condition, count(c.condition_name) as num, avg(toFloat(d.rating)) as rating
where num >20
return condition, num, rating
order by rating desc
```

For a given condition, are there numerous negatively-regarded drugs on the market?

Initial exploration for all drugs:

```
# Return the top 10 conditions with the lowest average rating
match (c:Condition)-[r:TREATS]-(d:Drug)
return c.condition_name as condition, count(c.condition_name) as num, avg(toFloat(d.rating)) as rating
order by rating
limit 10
```

Further exploration of peripheral t-cell lympho:

```
# Return graph with all the drugs used to treat peripheral t-cell lympho
```

```
match (c:Condition{condition_name:"peripheral t-cell lympho"})-[r:TREATS]-(d:Drug)
return c, d
```

```
# Return all the drug names and ratings for peripheral t-cell lympho
match (c:Condition{condition_name:"peripheral t-cell lympho"})-[r:TREATS]-(d:Drug)
return  d.drug_name as drug, d.rating as rating
```

Generalize for the market by only looking into conditions that have more than 20 associated drugs:

```
# Return the lowest average drug ratings and number of associated drugs for conditions
match (c:Condition)-[r:TREATS]-(d:Drug)
with  c.condition_name as condition, count(c.condition_name) as num, avg(toFloat(d.rating)) as
rating
where num >20
return condition, num, rating
order by rating
```

```
# Return the average sentiment score and number of associated drugs for conditions
match (c:Condition)-[r:TREATS]-(d:Drug)
with  c.condition_name as condition, count(c.condition_name) as num,
avg(toFloat(d.sentiment_score)) as sentiment
where num >20
return condition, num, sentiment
order by sentiment
```

Are there drugs that are able to treat multiple conditions? Which drug excels in doing so?

```
# Return drugs that are able to treat the most number of conditions and their average rating
match (c:Condition)-[r:TREATS]-(d:Drug)
with d.drug_name as drug, count(d.drug_name) as num, avg(toFloat(d.rating)) as rating
return drug, num, rating
order by num desc, rating desc
```

```
# Return graph with all the conditions prednisone treats
match (d:Drug {drug_name: 'prednisone'})-[r:TREATS]-(c:Condition)
return d,c
```

Are there conditions that can be treated by different drugs? Which drugs work the best?

```
# Return conditions that can be treated by multiple drugs and the average drug rating for each condition
match (c:Condition)-[r:TREATS]-(d:Drug)
with  c.condition_name as condition, count(c.condition_name) as num, avg(toFloat(d.rating)) as rating
return condition, num, rating
order by num desc, rating desc

# Return average effectiveness for conditions that can be treated by multiple drugs
match (c:Condition)-[r:TREATS]-(d:Drug)
with  c.condition_name as condition, count(c.condition_name) as num,
avg(toFloat(d.effectiveness)) as effectiveness
return condition, num, effectiveness
order by num desc, effectiveness desc

# Return graph with all the drugs that treat pain
match (c:Condition {condition_name: 'pain'})-[r:TREATS]-(d:Drug)
return d,c
```

Which gene is most commonly found in various conditions?

```
# Return the top genes and the number of associated conditions with that gene
match (c:Condition)-[r:ASSOCIATED]-(g:Gene)
return g.gene_name as gene, count(g.gene_name) as num
order by num desc

# Return graph with all the conditions and drugs associated with CD14
match (g:Gene {gene_name: "CD14"})--(c:Condition)--(d:Drug)
return g,c,d
```

For drugs that treat multiple conditions, are there commonly expressed genes for said conditions?

```
# Return genes that are associated with multiple conditions that prednisone treats
match
(g:Gene)-[r:ASSOCIATED]-(c:Condition)-[t:TREATS]-(d:Drug{drug_name:'prednisone'})
WITH g.gene_name as gene, count(g.gene_name) as num
WHERE num > 1
return gene, num
order by num desc

# Return graph with the common genes and associated conditions for prednisone
```

```
 match
(g:Gene)-[r:ASSOCIATED]-(c:Condition)-[t:TREATS]-(d:Drug{drug_name:'prednisone'})
where g.gene_name = "IL10" or g.gene_name="TNF"or g.gene_name="CD14" or
g.gene_name="HLA-DRB1" or g.gene_name="HLA"
return g,c,d
```

For conditions with poorly rated drug treatments, are there common genetic factors?

```
# Return associated genes and conditions for drugs with an average rating less than 5.0
match (g:Gene)--(c:Condition)-[r:TREATS]-(d:Drug)
with  c, g, d, count(c.condition_name) as num, avg(toFloat(d.rating)) as rating
where rating < 5.0
return g, c, d
```