



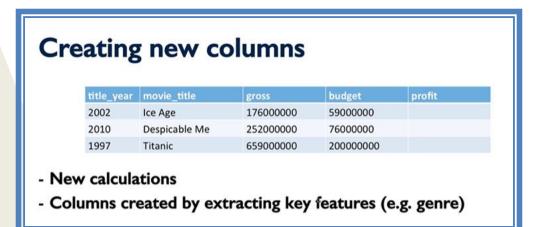
# RECAP....

```
select title_year, movie_title, category, `Won?`
from t1, t2
where t1.movie_title = t2.Nominee
and t1.title_year = t2.Year
and `Won?` = "YES"
order by title_year desc
```





## Creating new Columns...







## Missing / Erroneous Data....

#### Missing/erroneous data

title_year	movie_title	gross	budget
2006	The Host	2201412	1220000000
2010	Despicable Me	252000000	76000000
2003	Girl With A Pearl Earring	11634362	

Real-world data is often messy.

You need to think about how to deal with it.

Check integrity, fix errors, etc.

Simply removing bad data could also be problematic.

### Social Media Data....

#### Social media data

- Extract metadata
- Parse and process text
- Extract information from images/videos
- Draw connections between different posts
- Mine the "social graph" to build and extract data from network.







# THANKS!

Any questions?