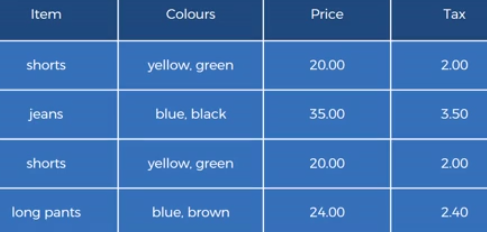
***Database Design***

*Normalization*

* **Normalization –** process of organizing columns/attributes + tables/relations or an RDB to reduce data redundancy and improve data integrity
* Means to better structure DB’s to remove the potential for any UPDATE/INSERT/DELETE anomalies + speed up the process of how the DB is utilized



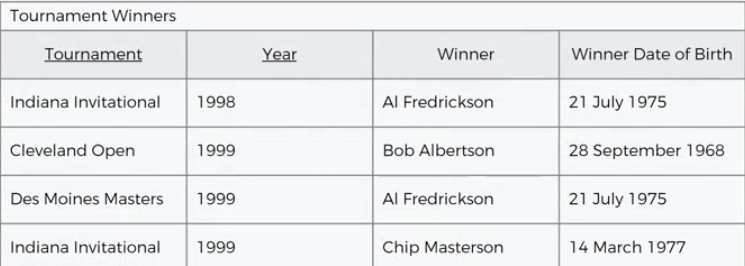
* See 2 values in each cell for Colours, + we just want 1 value for an attribute
* Could see 2 identical rows (shorts) 🡪 2nd row adds no value 🡪 don’t want
* Let’s say color doesn’t impact price in any way + we want to manually add a new row w/ purple jeans
* Should be same price + tax as 1st row w/ jeans, b/c the new color doesn’t affect them
* Now we’re duplicating price + tax info for jeans 🡪 doing unnecessary work to get info that’s already present + wasting time + resources to write it again
* Also raises potential for error if we manually input the wrong price + tax for jeans
* DB Design is important for DS b/c by learning it, we will be able to:
* *Understand HOW + WHY* a DB we’re working w/ was designed in a certain way
* Understand how to navigate around the DB structure quicker
* See + understand potential vulnerabilities + flaws in the DB + data
* Communicate effectively w/ data custodians/stewards/engineers/architects/engineers

*Prime + Non-Prime Attributes*

* Reminder: a PK is a type of candidate key that is chosen to uniquely ID a row



* Employee ID = PK (an employee can only work in 1 dept. = assumption)
* **Prime attribute =** all columns that are part of a candidate key 🡪 Employee ID
* **Non-prime attributes** = columns that are NOT a part of ANY candidate key 🡪 Name, Dept. ID, Dept. name



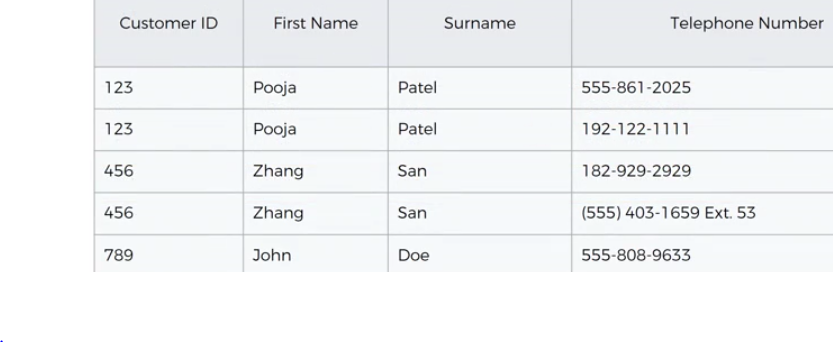
* Candidate Key = Tournament + Year 🡪 prime attributes
* Non-prim attributes = winner + winner DOB



* Prime attributes (candidate key) = Manufacturer + Model AND Model Full Name (2 CK’s)
* Model Full Name does not prevent Manufacturer + Model from being a CK on their own
* Non-prime attribute 🡪 Manufacturer Country

*First Normal Form (1NF)*

* 1NF 🡪 table that doesn’t contain duplicate rows + every cell contains only 1 value
* How to fix 🡪 remove the duplicate row or split out a cell w/ 2 values into 2 rows



* Pneumonic phrase to remember normal forms: **Every non-prime attribute must provide a fact about the key, the whole key, and nothing but the key**
* About the key = 1NF, whole key = 2NF, nothing but the key = 3NF
* “Doesn’t contain duplicate rows” = table has a key