**DATA PREPROCESSING**

* Why Do we need to Preprocess data?
  + Data is dirty (Incomplete, Noisy, Inconsistent)
* **Graphical user interface, text, application

  Description automatically generated**
* **Graphical user interface, text, application, email

  Description automatically generated**
* **Graphical user interface

  Description automatically generated with medium confidence**
* **Graphical user interface, text, application, email

  Description automatically generated**
* **Text

  Description automatically generated**
* **Text

  Description automatically generated**
* **SORTING Into Equal-depth bins**
* **Bin 1: 13, 15 ,16**
* **Bin 2: 16, 19, 20**
* **Bin 3: 20, 21, 22**
* **Bin 4: 22, 25, 25**
* **Bin 5: 25, 25, 30**
* **Bin 6: 33, 33, 35**
* **Bin 7: 35, 35, 35**
* **Bin 8: 36, 52, 70**

**SMOOTH BY BIN MEANS:**

**B1: 15, 15, 15**

**B2: 18, 18 ,18**

**B3: 21, 21, 21**

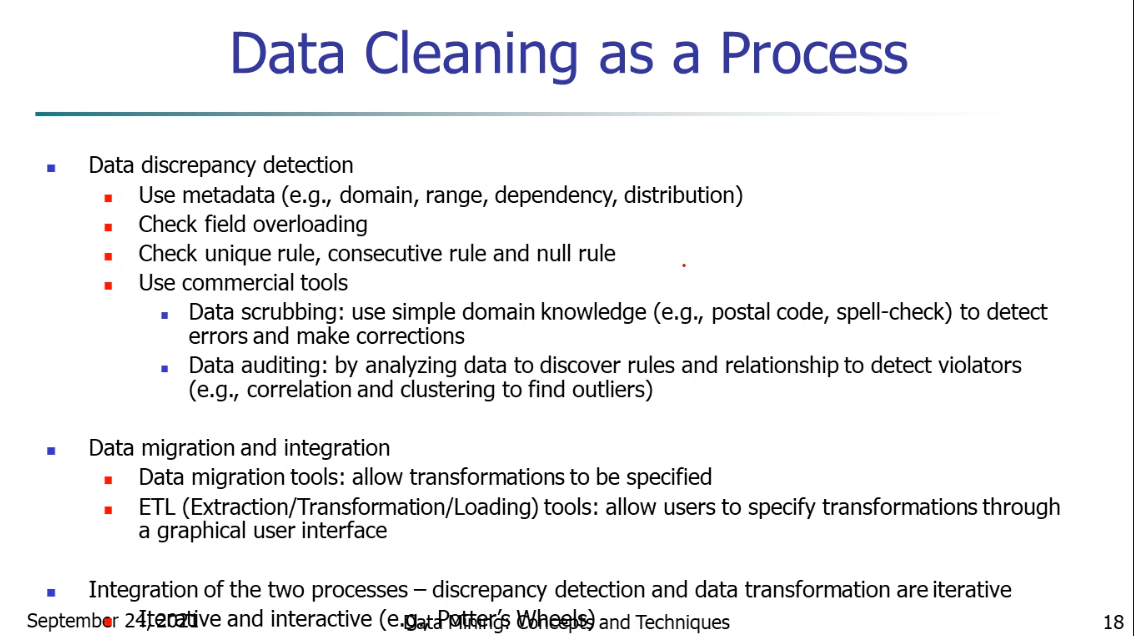
**B4: 24, 24, 24**

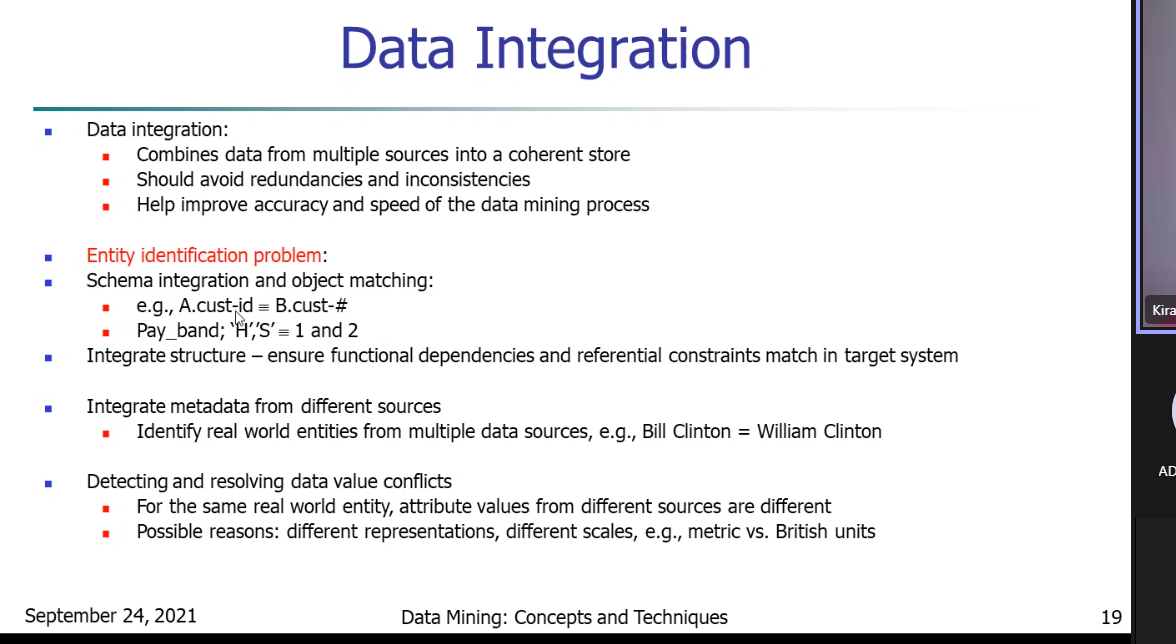
**B5: 27, 27, 27**

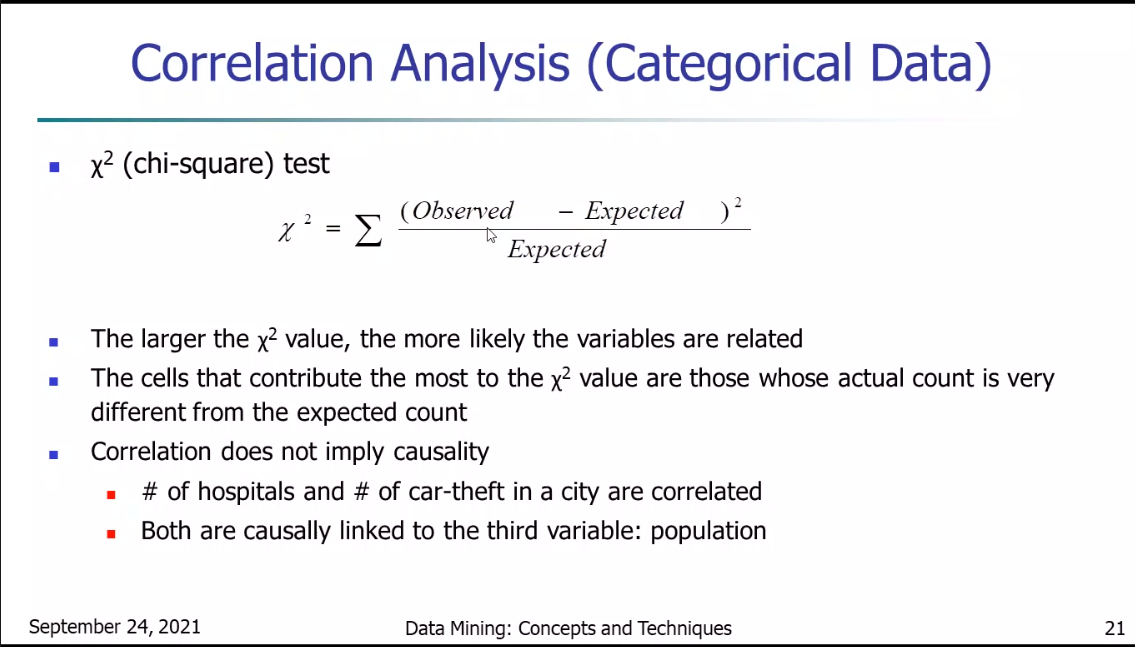
**B6: 34, 34, 34**

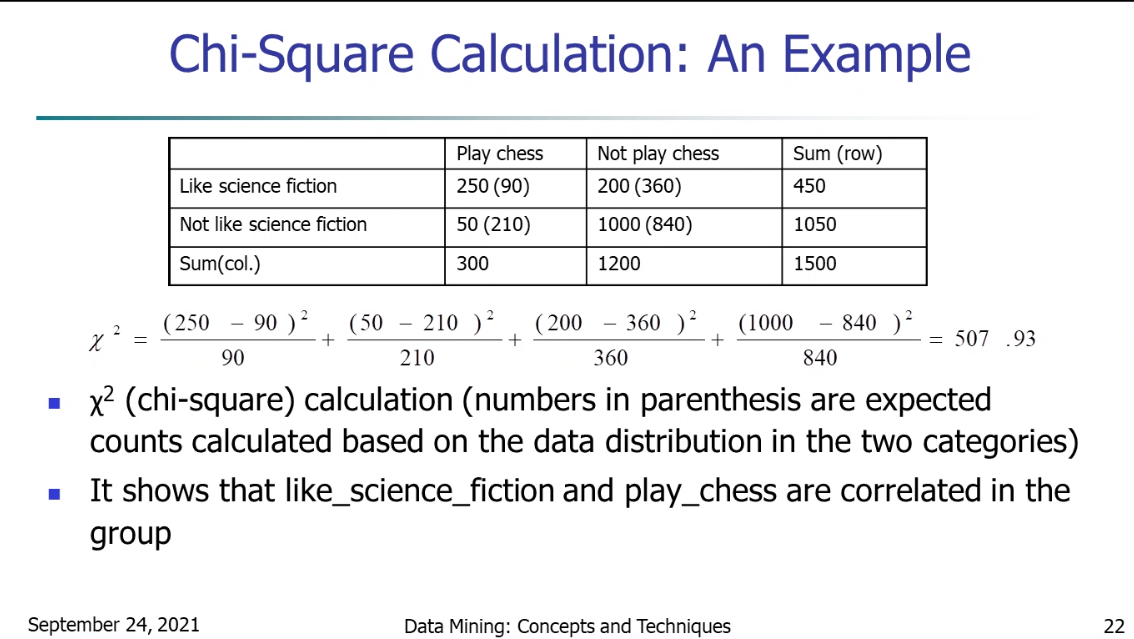
**B7: 35, 35, 35**

**B8: 53, 53, 53**

****

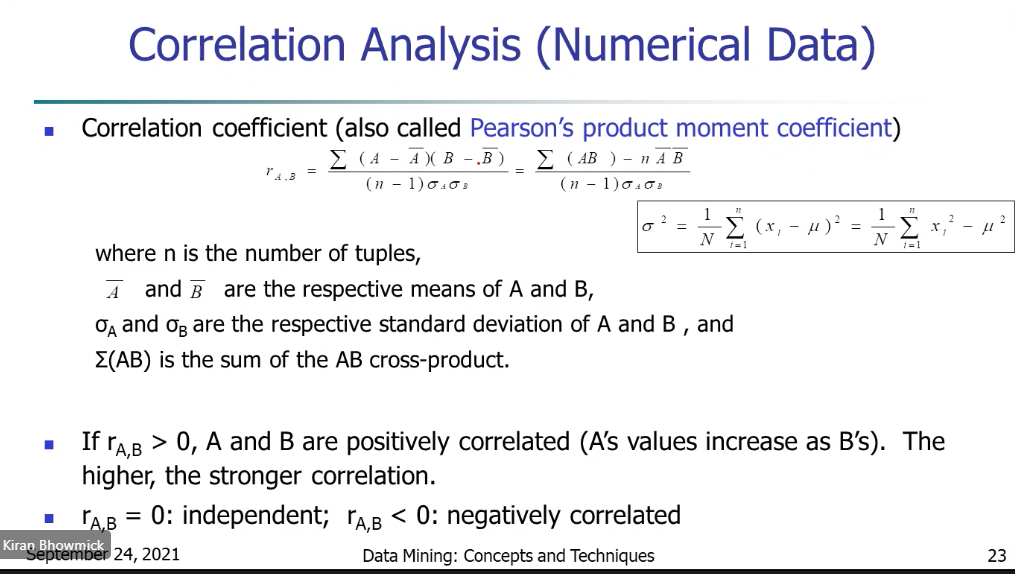
****

****

****

**Degree of Freedom = Number of attributes**

**Look for 95% probability of the DOF**

****