MIS 304: Using and Managing Information Systems

Assignment 2

Wenli Zhang

Summer II 2016

Errata

Assignment 2
Typo on P10:

(3) Fill in the column for "isMagazineSubscriber". The purpose is to know if the sampled individual subscribes to a magazine. This is used in answering Question_2. A formula you can use is:

Magazines

=IF(ISBLANK([@[Books / Magazines]]),"Unknown",IF(ISERROR(SEARCH("Magazine Subscriber",[@[Books / Magazines]])),"Non_Subscriber","Magazine Subscriber"))

	A	B	C	D	E	F	G	H	I	J	K	L
1	SampleNo	Last Na	Street	Street	City	State	Zip Cod	Estimated Ho	MidEHI	Estimate	Books / B	isMagazineSubscri
2	1	CERVANTI	Speedway	Blvd	Tucson	AZ	85719	$100,000 - $124,9	112500	$100,000 - $	Books & M	Magazine Subscriber

Goal

- Data preparation
- Excel functions
- Excel Pivot table

Data Preparation

- Today's **real-world** databases are highly susceptible to noisy, missing, and inconsistent data
- Reasons:
  - Huge size (often several gigabytes or more)
  - Machine and human error
  - Likely origin from multiple sources
- **Low-quality data will lead to low-quality analyzing results**

Data Preparation: Data Preprocessing & Data Cleaning

- **Data preprocessing:** transforming raw data into an understandable format
- **Data Cleaning**: detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database
- Data pre-processing and cleansing can take up to **80%** of the total analysis task time

Excel functions

- Why we need functions ?
- **DRY**: **don't repeat yourself !!!**

- **Reasons:**
  - **Productivity**
  - **Complexity**

Excel functions

- Build a **Chihuahua** wash machine:
  1. Brush your **Chihuahua** before a bath
  2. Use lukewarm water
  3. Talk to your **Chihuahua** in a calm voice
  4. Use **Chihuahua** shampoo
  5. Rinse well
  6. Air-dry
  7. Reward your **Chihuahua**

- Build a **Husky** wash machine:
  1. Brush your **Husky** before a bath
  2. Use lukewarm water
  3. Talk to your **Husky** in a calm voice
  4. Use **Husky** shampoo
  5. Rinse well
  6. Air-dry
  7. Reward your **Husky**

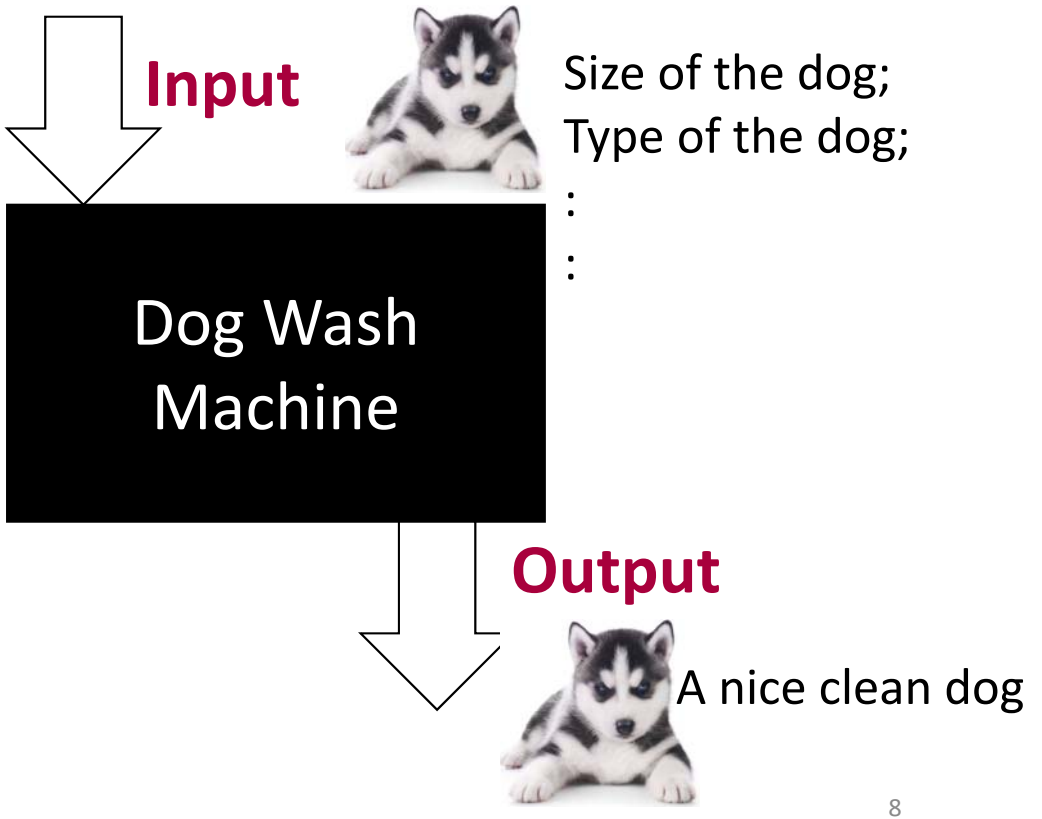
Excel functions

Build a Dog wash machine:

1. Brush your **Dog** before a bath.
2. Use lukewarm water
3. Talk to your **Dog** in a calm voice.
4. Use **Dog** shampoo.
5. Rinse well.
6. Air-dry.
7. Reward your **Dog**.

In Excel:

Functions are predefined **formulas** and are already available in Excel.
A **formula** is an **expression** which calculates the value of a cell.

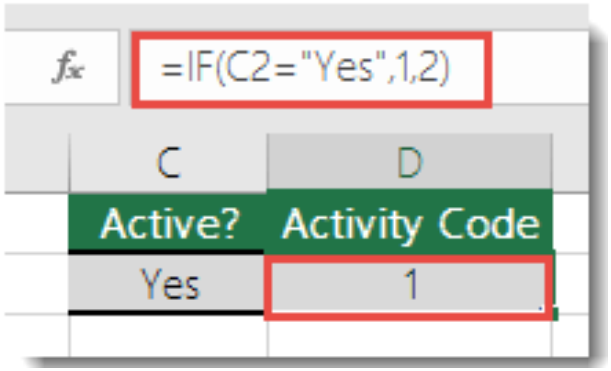


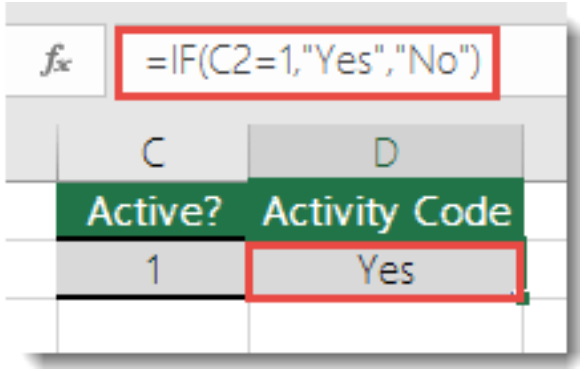
Excel Functions

Example: IF function

IF(logical_test, [value_if_true], [value_if_false])

IF(Something is True, then do something, otherwise do something else)





Excel Functions

Best way to learn Excel functions

https://support.office.com/en-us/article/Excel-functions-alphabetical-b3944572-255d-4efb-bb96-c6d90033e188

Excel Functions

- Nested IF function
- Example: Assignment 2: P10

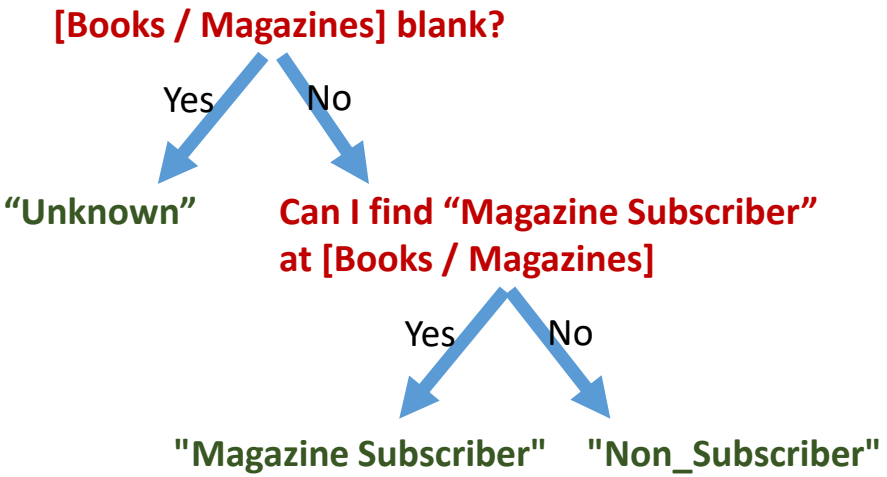
- Fill in the column for “isMagazineSubscriber”.
- The **purpose** is to know if the sampled individual subscribes to a magazine.

Books / Magazines	isMagazineSubscriber
Books & Magazines, Magazine Subscribers	
Book Buyers, Books & Magazines, Magazine Subscribers	
Books & Magazines, Magazine Subscribers	
349,999	
Books & Magazines	
Book Buyers, Books & Magazines, Comic Book Readers, Magazine Subscribers	
174,999	
Books & Magazines, Magazine Subscribers	
299,999	
Books & Magazines, Magazine Subscribers	
Books & Magazines, Magazine Subscribers	
Books & Magazines, Magazine Subscribers	
124,999	
Books & Magazines, Magazine Subscribers	
Books & Magazines, Magazine Subscribers	
9,999	
Book Buyers, Books & Magazines, Magazine Subscribers	
Books & Magazines, Magazine Subscribers	

IF(Something is True, then do something, otherwise do something else)

=**IF**(**ISBLANK**([@[Books / Magazines]]),"Unknown",

IF(**ISERROR**(**SEARCH**("Magazine Subscriber",[@[Books / Magazines]])),"Non_Subscriber", "Magazine Subscriber"))



Excel Pivot table

- One of Excel's most powerful features
- **Data summarization** tool
- Allows you to **extract** the significance from a large, detailed data set

Excel Pivot table

Best way to learn Excel Pivot table

Data types

- **Quantitative data** (continuous data):
  - The values can change continuously
  - Anything you can measure or count is quantitative
  - E.G, weight, price, profits, counts, etc
- **Categorical data**:
  - Data be divided into groups
  - You typically can list a small number of categories
  - E.G, Product type, gender, age group, etc.

Data types

- Most data sets contain **both types** of data.
- It's actually quite difficult to summarize your data that is purely quantitative or purely categorical

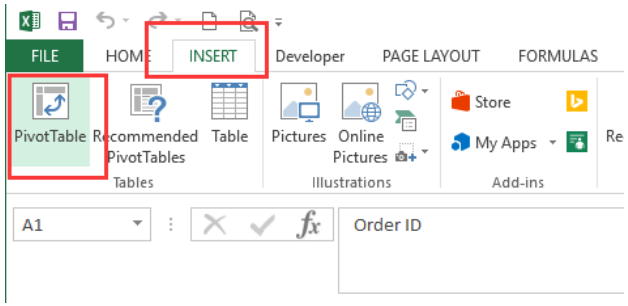
Excel Pivot table

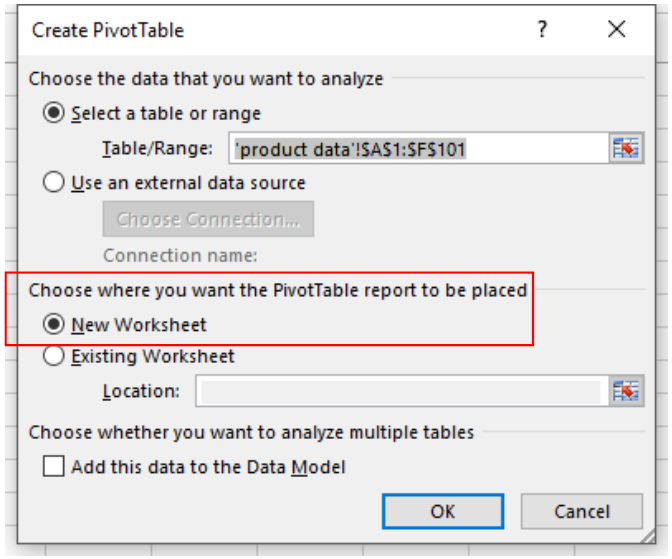
- Open ***in_class_practice.xlsx*** (Blackboard)
- Which ones are **numerical**, and which ones are **categorical**?

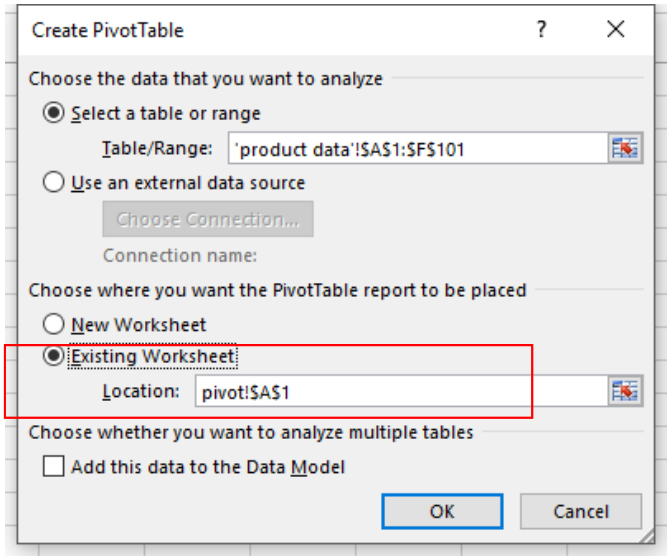
Order ID	Product	Category	Amount	Country
1	Carrots	Vegetables	$4,270	United States
2	Broccoli	Vegetables	$8,239	United Kingdom
3	Banana	Fruit	$617	United States
4	Banana	Fruit	$8,384	Canada

Insert a Pivot Table

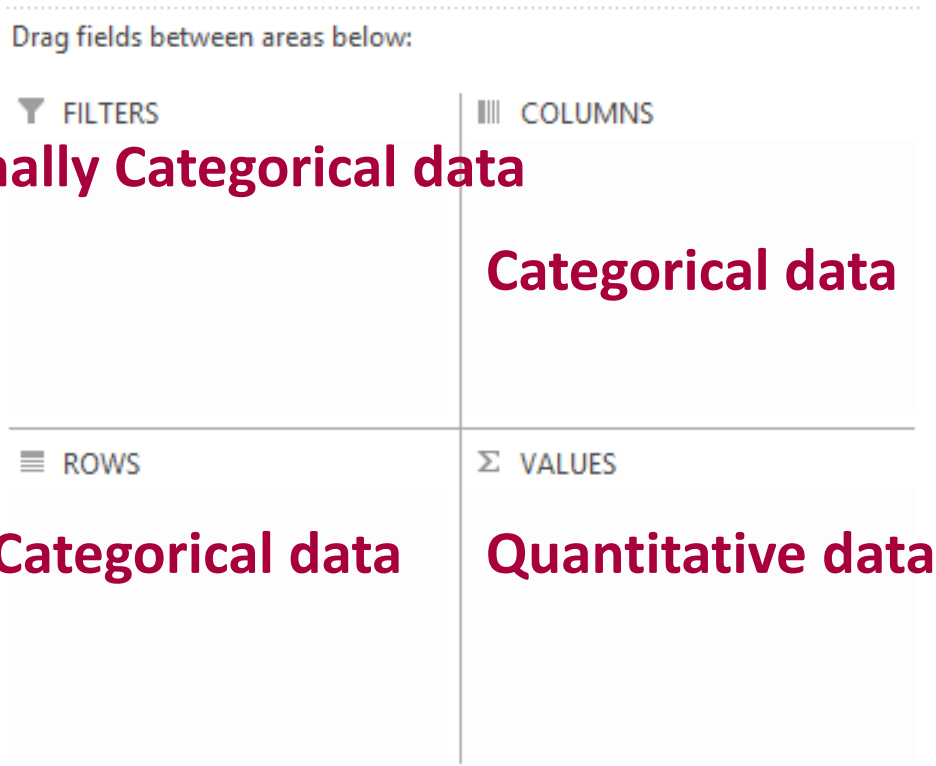
1. Click any **single cell inside** the data set.
2. On the **Insert** tab, click **PivotTable**.







Normally



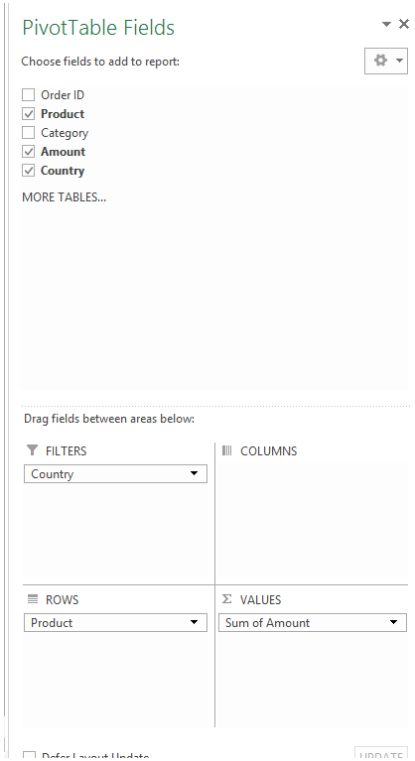
Normally Categorical data

Categorical data

Categorical data

Quantitative data (Could be Categorical)

Drag fields

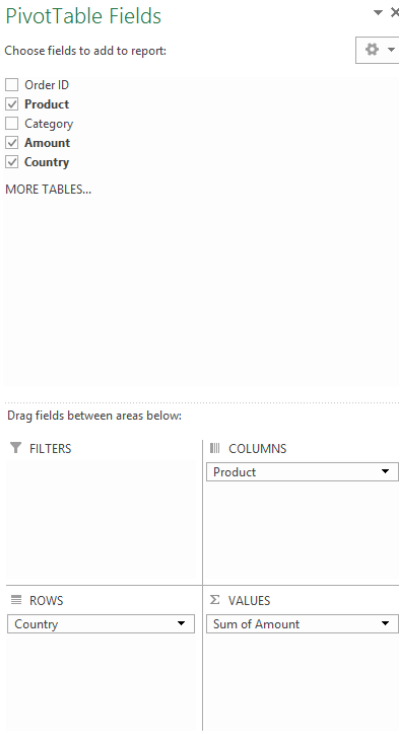


	A	B
1	Country	(All)
2		
3	Row Labels	Sum of Amount
4	Apple	93057
5	Banana	150702
6	Beans	32590
7	Broccoli	59557
8	Carrots	62745
9	Mango	38662
10	Orange	46638
11	Grand Total	483951

Drag:

1. **Product** Field to the Row Labels area
2. **Amount** Field to the Values area
3. **Country** Field to the Filter area

Two-dimensional Pivot Table



Sum of Amount	Column Labels							
Row Labels	Apple	Banana	Beans	Broccoli	Carrots	Mango	Orange	Grand Total
Australia	18874	17677	1002	9062	5154	3663	2493	57925
Canada	14558	18168		12055			9939	54720
France	43861	16751		5341		7388	2256	75597
Germany		13754	20866	10188	8357	8775	8887	70827
New Zealand	5820	22531						28351
United Kingdom		20758	3559	15899	31218	5600	8141	85175
United States	9944	41063	7163	7012	18016	13236	14922	111356
Grand Total	**93057**	**150702**	**32590**	**59557**	**62745**	**38662**	**46638**	**483951**

1. **Country** Field to the Row Labels area.
2. **Product** Field to the Column Labels area.
3. **Amount** Field to the Values area.