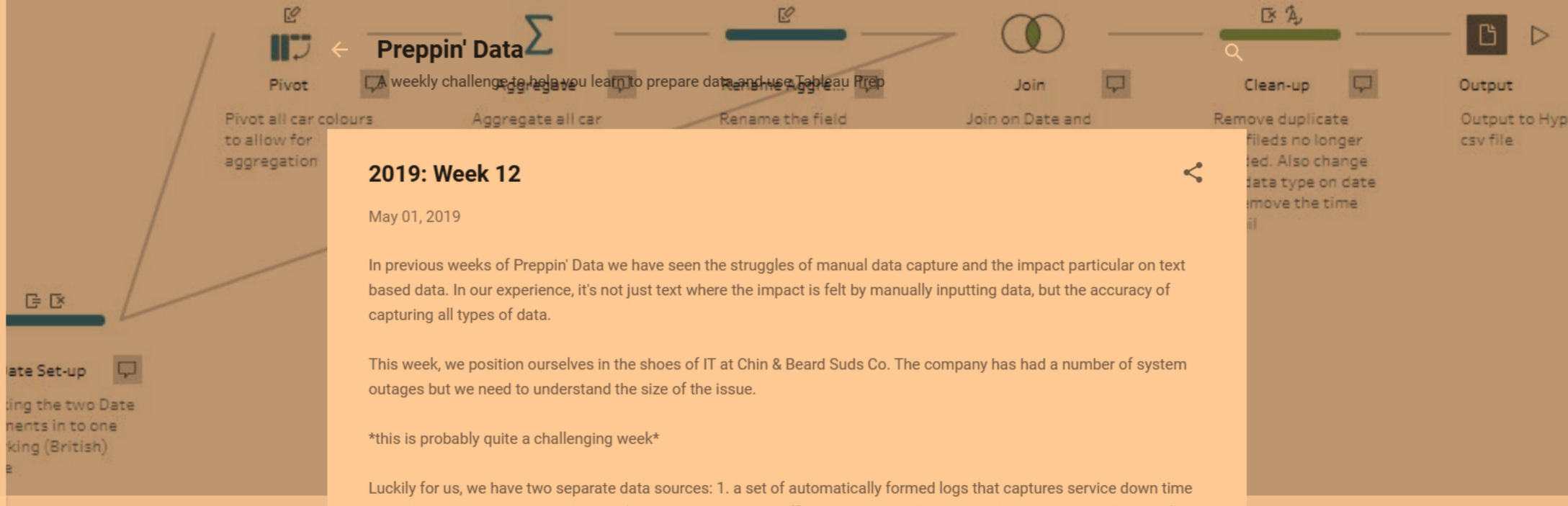


Contributors

CARL ALLCHIN

JENNY MARTIN

JONATHAN ALLENBY

TOM PROWSE

Followers

フォロワー（203 人） 次へ

フォローする

Archive

Labels

Report Abuse

2019: Week 12

May 01, 2019

In previous weeks of Preppin' Data we have seen the struggles of manual data capture and the impact particular on text based data. In our experience, it's not just text where the impact is felt by manually inputting data, but the accuracy of capturing all types of data.

This week, we position ourselves in the shoes of IT at Chin & Beard Suds Co. The company has had a number of system outages but we need to understand the size of the issue.

this is probably quite a challenging week

Luckily for us, we have two separate data sources: 1. a set of automatically formed logs that captures service down time with a precise timestamp; 2. a manual spreadsheet where staff can capture issues they are having with systems. Sadly, for the latter data source, the data is captured without the same accuracy as the automatic logs. With the automatic logs, have we got all the data though? Help us clean the data and make clear how much downtime we are suffering from, in which system and what error causes the biggest % of downtime.

Requirements

Start Date / Time	End Date / Time	System	Error
13/04/2019 07:55:23	13/04/2019 08:23:12	Sales	Disc full
13/04/2018 09:03:22	15/04/2018 09:03:21	Stock	Planned Outage
13/05/2018 09:03:22	16/05/2018 09:03:21	Stock	Planned Outage
13/06/2018 09:03:22	15/06/2018 09:03:21	Stock	Planned Outage
12/07/2018 23:03:22	13/07/2018 07:00:03	Stock	Planned Outage
12/08/2018 23:03:22	13/08/2018 07:00:03	Stock	Planned Outage
19/08/2018 07:55:20	19/08/2018 08:31:55	Sales	Unknown disc location
12/09/2018 23:03:22	13/09/2018 07:31:02	Stock	Planned Outage
19/09/2018 14:57:00	19/09/2018 15:02:20	Sales	Unknown disc location

Auto input

Start Date	Start Time	End Date	End Time	System	Error
13/04/2018	09:00:00	15/04/2018	09:00:00	Stock	Planed Outage
13/05/2018	09:00:00	15/05/2018	09:00:00	Stock	Planed Outage
13/06/2018	09:00:00	15/06/2018	09:00:00	Stock	Planed Outage
13/07/2018	09:00:00	15/07/2018	09:00:00	Stock	Planed Outage
13/08/2018	09:00:00	15/08/2018	09:00:00	Stock	Planed Outage
19/09/2018	15:00:00	19/09/2018	15:15:00	Sales	Sue tripped on a cable

Manual input

- Input the data
- Make the manual date and time a date / time field
- Bring the datasets together in a manner that removes the duplicate records from the manually captured data set
  - Duplicates are determined by a system being down at the same time but recorded both automatically and manually. The automatic data should always overwrite the manual data.
- Workout the duration (in seconds) of the error
- Understand the '% of downtime' per system in hours

Output

% of system downtime	Total Downtime in Hours	Downtime in Hours	Error Source	Error	Start Date / Time	End Date / Time	System
0.4	1.2	0.5	Automatic Error log	Disc full	13/04/2019 07:55	13/04/2019 08:23	Sales
0.17	288.3	48	Automatic Error log	Planned Outage	13/04/2018 09:03	15/04/2018 09:03	Stock
0.25	288.3	72	Automatic Error log	Planned Outage	13/05/2018 09:03	16/05/2018 09:03	Stock
0.17	288.3	48	Automatic Error log	Planned Outage	13/06/2018 09:03	15/06/2018 09:03	Stock
0.03	288.3	7.9	Automatic Error log	Planned Outage	12/07/2018 23:03	13/07/2018 07:00	Stock
0.03	288.3	7.9	Automatic Error log	Planned Outage	12/08/2018 23:03	13/08/2018 07:00	Stock
0.52	1.2	0.6	Automatic Error log	Unknown disc location	19/08/2018 07:55	19/08/2018 08:31	Sales
0.01	288.3	8.5	Automatic Error log	Planned Outage	12/09/2018 23:03	13/09/2018 07:31	Stock
0.08	1.2	0.1	Automatic Error log	Unknown disc location	19/09/2018 14:57	19/09/2018 15:02	Sales
0.17	288.3	48	Manual capture error list	Planned Outage	13/06/2018 09:00	15/06/2018 09:00	Stock
0.17	288.3	48	Manual capture error list	Planned Outage	13/07/2018 09:00	15/07/2018 09:00	Stock

- 8 columns
- 11 rows (12 including headers)
- No nulls

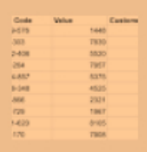
For comparison, here's our output file. Don't to forget to fill in our participation tracker!

Aggregation Date Calculations Join Union

Popular posts from this blog

2023: Week 1 The Data Source Bank

January 04, 2023



Created by: Carl Allchin Welcome to a New Year of Preppin' Data. These are weekly exercises to help you learn and develop data preparation skills. We publish the challenges on a Wednesday and share a solution the following Tuesday. You can take the challenges whenever you want and we love to see y ...

READ MORE

2021: Week 22 - Answer Smash

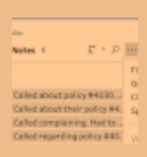
June 02, 2021

Challenge By: Jenny Martin Recently, my family and I have become quite invested in the TV quiz show Richard Osman's House of Games . The final round is always a round called Answer Smash. In this round you have a picture and question and you have to "smash" the name of the picture with the ansv ...

READ MORE

How to...Handle Free Text

January 24, 2020



Free isn't always a good thing. In data, Free text is the example to state when proving that statements correct. However, lots of benefit can be gained from understanding data that has been entered in Free Text fields. What do we mean by Free Text? Free Text is the string based data that comes from allowi ...

READ MORE

Powered by Blogger