Black friday  
Deals of a Lifetime

November 17, 2018

## Project Background and Description

|  |  |
| --- | --- |
|  | We selected a timely topic, Black Friday shopping, and then collected data about day-after-Thanksgiving sales and America’s shopping habits. Black Friday can be a great day to score deals for yourself or to cross off gifts on your holiday shopping list. Alternatively, you could be one of the many people involved in an incident either as the victim or perpetrator of a crime. Unsurprisingly, more thefts are reported on Black Friday than any other day of the year. Surprisingly, since 2006 there have been 111 injuries and 10 deaths making national headlines on Black Friday. |

## Database Design and Tables

|  |  |
| --- | --- |
|  | Our **Black Friday** database table names include: Incidents, and Black Friday Deals. All Black Friday data is stored in these tables. |

* The “**Black Friday Deals**” table contains categories related to the specific stores, types of deals that store offered, and the specific sale or discount.
* The “**Incidents**” table contains data related to outrageous occurrences – assaults, deaths, injuries – that made national news during Black Friday shopping, arranged by stores and cities.

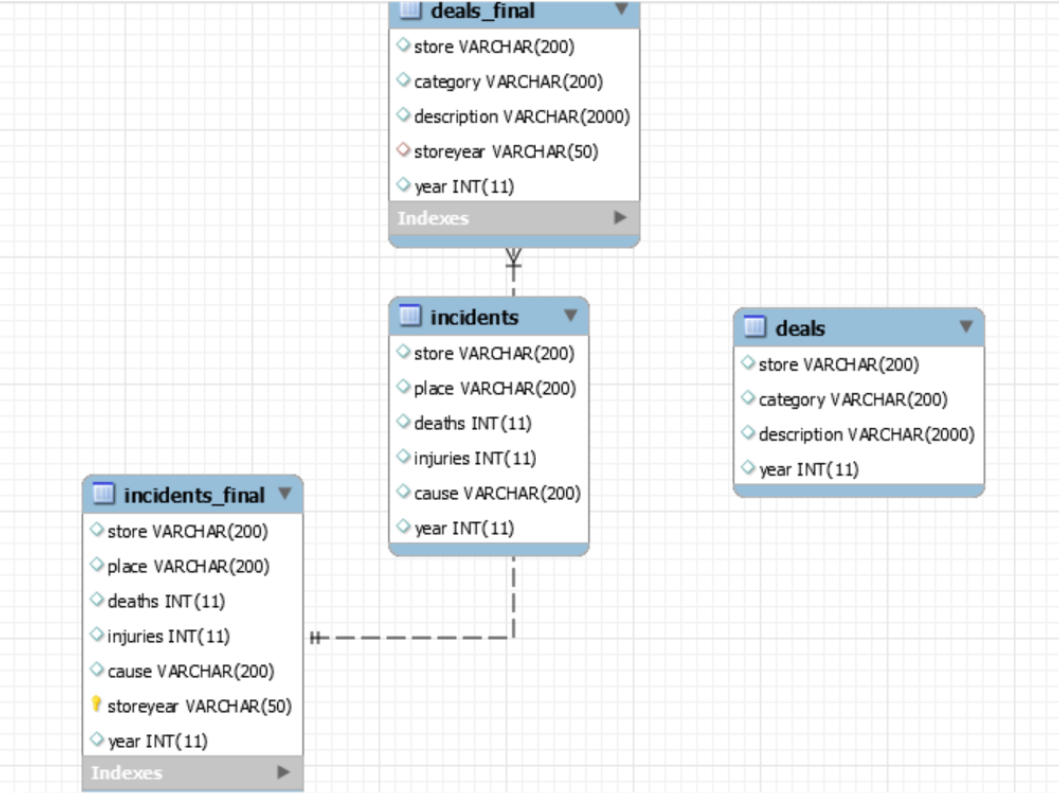
## Data Sources and Cleaning the Data

|  |  |
| --- | --- |
|  | Data sources are CSVs from Socrata and Kaggle. The data was extracted from CSV via Python Pandas and Jupyter Notebook. Tables were created in MySQL and then filled with the data after it was organized in Pandas. |

* Pandas helped clean and join the documents. We ran into quite a few errors with loading the data. The CSVs had to be cleaned of a few blank cells, null values, and data links. They also needed to be parsed into the correct delimited format. We had to remove a couple of rouge commas that were throwing errors in Jupyter Notebook. Ultimately, we slimmed down our datasets to be quite small after lots of trial and error.
* One challenge was that there wasn’t enough unique data to assign foreign a key. We had to assign greater detail to some of the stores.
* The two CSVs we used each had different spelling for names such as: Walmart vs. Wal-Mart, Toys ‘R’ Us vs. Toys R Us. That was cleaned up prior to importing.

## Database Schema

|  |  |
| --- | --- |
|  | Two final tables, Incidents\_final and Deals\_final, are connected on the store, year. Each were taken from deals and incidents. |



## References

|  |  |
| --- | --- |
|  | [*https://www.thestreet.com/story/14399311/1/how-to-avoid-a-fight-on-black-friday.html*](https://www.thestreet.com/story/14399311/1/how-to-avoid-a-fight-on-black-friday.html)  [*https://en.wikipedia.org/wiki/Black\_Friday\_(shopping)*](https://en.wikipedia.org/wiki/Black_Friday_(shopping)) |

