Popular Halloween Candy

October, 28 2020

Team C:

Sara Behnke

Cory Potter

Han Zhao

❖ Introduction

This topic came at just the right time! We want to find out what the most popular Halloween candies are. This team will look at a survey of Halloween candy ranking and best selling candy data from Walgreen.com. The analysis will tell us:
➢ Most popular Halloween candy by survey
➢ Most popular Halloween candy by sales
➢ Most popular Halloween candy by review ratings

❖ Data source
➢ Candy-data.csv(Kaggle.com) includes attributes for each candy along with its ranking. For binary variables, 1 means yes, 0 means no. The data contains the following fields:
■ chocolate: Does it contain chocolate?
■ fruity: Is it fruit flavored?
■ caramel: Is there caramel in the candy?
■ peanutalmondy: Does it contain peanuts, peanut butter or almonds?
■ nougat: Does it contain nougat?
■ crispedricewafer: Does it contain crisped rice, wafers, or a cookie component?
■ hard: Is it a hard candy?
■ bar: Is it a candy bar?
■ pluribus: Is it one of many candies in a bag or box?
■ sugarpercent: The percentile of sugar it falls under within the data set.
■ pricepercent: The unit price percentile compared to the rest of the set.
■ winpercent: The overall win percentage according to 269,000 matchups.
➢ Walgreen's website:
https://www.walgreens.com/store/c/halloween-candy/ID=520931-tier3 (page is filtered by top sellers)

❖ Data Extraction: we performed web scraping for Walgreen's Halloween candy sales & rating data with beautiful soup.
➢ Steps:
■ Create path to read Walgreen HTML file.
■ Create beautiful soup object.
■ Extract all candy listings on the current page (72).
■ Create a list of dictionaries to return brand name, description of candy and its rating.
■ Use try and except block to convert the non-review/rating results to NaN
➢ Challenges:
■ When we first ran our scrape file, Walgreens was only returning 8 results and we were unable to figure out why that was.
➢ Solution:
■ We decided to copy the entire page's outer html and save that to our repository. We used os to read in the html file and then we were able to perform our scraping.

❖ Data Transformation
  ➢ Walgreen data transformation steps:
    ■ Create a dataframe with the list of dictionaries.
    ■ Drop any results that did not have a rating.
    ■ Break rating and reviews into two separate columns from the rating column.

	brand	description	rating
0	PayDay	Snack Size Peanut Caramel Bars Peanut Caramel	4.6 out of 5 total 12 reviews
1	Snickers	Peanut Butter Squared Fun Size Candy	5.0 out of 5 total 1 reviews
2	Tootsie Roll	Caramel Apple Pops Green Apple	5.0 out of 5 total 6 reviews
3	Milky Way	Chocolate Candy Bar Fun Size	4.9 out of 5 total 44 reviews
4	Reese's	Snack Size Peanut Butter Cups	5.0 out of 5 total 3 reviews

	brand	description	rating	individual_rating	review
0	PayDay	Snack Size Peanut Caramel Bars Peanut Caramel	4.6 out of 5 total 12 reviews	4.6	12.0
1	Snickers	Peanut Butter Squared Fun Size Candy	5.0 out of 5 total 1 reviews	5.0	1.0
2	Tootsie Roll	Caramel Apple Pops Green Apple	5.0 out of 5 total 6 reviews	5.0	6.0
3	Milky Way	Chocolate Candy Bar Fun Size	4.9 out of 5 total 44 reviews	4.9	44.0
4	Reese's	Snack Size Peanut Butter Cups	5.0 out of 5 total 3 reviews	5.0	3.0

    ■ Group by the brands to see the average brand rating and reviews.

brand	Avg_Rating	Avg_Num_Reviews
3 Musketeers	4.600000	16.000000
Airheads	4.000000	2.000000
Almond Joy	4.100000	24.000000
Blow Pop	5.000000	2.000000
Dubble Bubble	3.800000	4.000000
Dum Dums	5.000000	2.000000
Hershey's	4.680000	35.400000

  ➢ Survey data transformation steps:
    ■ Remove special characters and replace with apostrophes.
    ■ Create a dataframe for chocolate candies.

- Create a dataframe for fruity candies.
- Create a dataframe of all other candy types.
- Create a new table that only displays binary results.
- Create a new table for percentages.

	competitorname	brand	winpercent	sugarpercent	pricepercent
52	Reese's Peanut Butter Cup	Reese's	84.180290	0.720	0.651
51	Reese's Miniatures	Reese's	81.866257	0.034	0.279
79	Twix	Twix	81.642914	0.546	0.906
28	Kit Kat	Kit Kat	76.768600	0.313	0.511
64	Snickers	Snickers	76.673782	0.546	0.651

❖ Load data to database
  ➢ Walgreen data load steps:
    ■ Create connection to postgresql.
    ■ Use pandas to load the scraped Walgreens dataframe into database.
      ● Verify table was loaded successfully.
    ■ Use pandas to load brand averages dataframe to database
      ● Verify table was loaded successfully.
  ➢ Survey data load steps:
    ■ Create connection to postgresql.
    ■ Use pandas to load binary_results dataframe into database.
      ● Verify table was loaded successfully.
    ■ Use pandas to load percentage dataframe to database
      ● Verify table was loaded successfully.

Final Screenshots of posgres tables:

```
1   select * from walgreens_brand_avg
2
```

Data Output   Explain   Messages   Notifications

	brand text	Avg_Rating double precision	Avg_Num_Reviews double precision
1	3 Musket…	4.6	16
2	Airheads	4	2
3	Almond J…	4.1	24
4	Blow Pop	5	2
5	Dubble B…	3.8	4
6	Dum Dums	5	2
7	Hershey's	4.68	35.4
8	Kathy Kaye	3	2
9	Kisses	5	1
10	Kit Kat	3.6	6.5

```
1   select * from survey_percentage_results
2
```

Data Output   Explain   Messages   Notifications

	competitorname text	brand text	winpercent double precision	sugarpercent double precision	pricepercent double precision
1	Reese's Peanut Butter …	Reese's	84.18029	0.72000003	0.65100002
2	Reese's Miniatures	Reese's	81.86625699999999	0.034000002	0.27900001
3	Twix	Twix	81.642914	0.546	0.90600002
4	Kit Kat	Kit Kat	76.7686	0.31299999	0.51099998
5	Snickers	Snickers	76.67378199999999	0.546	0.65100002
6	Reese's Pieces	Reese's	73.43499	0.40599999	0.65100002
7	Milky Way	Milky Way	73.099556	0.60399997	0.65100002
8	Reese's stuffed with pi…	Reese's	72.887901	0.98799998	0.65100002
9	Peanut butter M&MÕs	M&M's	71.46505	0.82499999	0.65100002
10	Nestle Butterfinger	Nestle	70.735641	0.60399997	0.76700002

```sql
1  select * from survey_binary_results
2
```

Data Output Explain Messages Notifications

	competitorname text	brand text	chocolate bigint	fruity bigint	caramel bigint	peanutyalmondy bigint	nougat bigint	crispedricewafer bigint	hard bigint	bar bigint	pluribus bigint
1	100 Grand	100 Grand	1	0	1	0	0	1	0	1	0
2	3 Musketeers	3 Musket...	1	0	0	0	1	0	0	1	0
3	One dime	One dime	0	0	0	0	0	0	0	0	0
4	One quarter	One quart...	0	0	0	0	0	0	0	0	0
5	Air Heads	Airheads	0	1	0	0	0	0	0	0	0
6	Almond Joy	Almond J...	1	0	0	1	0	0	0	1	0
7	Baby Ruth	Baby Ruth	1	0	1	1	1	0	0	1	0
8	Boston Baked Beans	Boston B...	0	0	0	1	0	0	0	0	1
9	Candy Corn	Brach's	0	0	0	0	0	0	0	0	1
10	Caramel Apple Pops	Tootsie R...	0	1	1	0	0	0	0	0	0

```sql
1  select * from survey_percentage_results
2
```

Data Output Explain Messages Notifications

	competitorname text	brand text	winpercent double precision	sugarpercent double precision	pricepercent double precision
1	Reese's Peanut Butter ...	Reese's	84.18029	0.72000003	0.65100002
2	Reese's Miniatures	Reese's	81.86625699999999	0.034000002	0.27900001
3	Twix	Twix	81.642914	0.546	0.90600002
4	Kit Kat	Kit Kat	76.7686	0.31299999	0.51099998
5	Snickers	Snickers	76.67378199999999	0.546	0.65100002
6	Reese's Pieces	Reese's	73.43499	0.40599999	0.65100002
7	Milky Way	Milky Way	73.099556	0.60399997	0.65100002
8	Reese's stuffed with pi...	Reese's	72.887901	0.98799998	0.65100002
9	Peanut butter M&MÕs	M&M's	71.46505	0.82499999	0.65100002
10	Nestle Butterfinger	Nestle	70.735641	0.60399997	0.76700002