

INTRODUCTION

Delta Faucet Company is the world's leading innovator in faucets, flush valves and related accessories.

Goal: Identify current and future consumer design/feature preferences for kitchen and bathroom faucets using social media and other websites. Based on the scope of the project, it is divided into three problem spaces –

Problem Space 1:

- Scan and summarize latest feature trends
- Identify market and design opportunity using Social Media sites (Twitter)

Problem Space 2:

- Use online social and other data to scrape product information
- Analyze the scraped data to prioritize existing or trending features/concepts

Problem Space 3:

- Analyze customer satisfaction data to identify areas of opportunities
- Categorize and prioritize the opportunities to improve customer service

METHODOLOGY

Problem Space 1: Our project is split into the following phases:



Figure 1.1: Project Workflow

Data Collection: Used Tweepy API through Python

- Utilized keywords, date, and location to filter
- Extracted username, images, number of likes and comments

Data Cleaning: Removed irrelevant tweets

METHODOLOGY

Data Analysis: Sentiment Analysis and engagement analytics to rank tweets based on popularity

Identify:

- Zero-shot text classification model
- Basic image classification model

Compile Results: Retrieved topics from the most popular tweets

Problem Space 2:

Web Scrapers:

- Utilized BeautifulSoup and Selenium
- Scraped product specifications for various retailer's websites e.g. – Amazon, Home Depot etc.
- Analyzed the data for popular trends

Automation of web scrapers:

- Scrapers feed system with data continuously
- Back-end connects database to system
- Front-end displays results to Delta Faucet employees

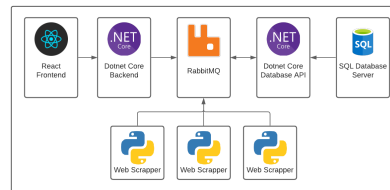


Figure 2.1: Diagram of Automation System

Problem Space 3:

- Utilized Python's BeautifulSoup and Selenium libraries
- Scikit-learn library for supervised classification.
- Built a dataset of scraped reviews.
- Built web app to display the dataset as well as some basic analytics and periodically scrape the web for additional reviews.

RESULTS

Problem Space 1:

Data Collection:

- Scraped 60+ tweets, with their text, images, username, location, number of likes, & number of retweets

	text	name	location	favorite	retweet
17	Hey friends that have a good sense for design/...	jessedriftwood1	NaN	39	1
15	@kareem_carr @arthur_affect I was in one of th...	anng27	Seattle, WA, USA	9	0
31	The organically shaped ZA #faucet matches the ...	TOTOUSA	NaN	7	3
39	We're here to say it and spray it. We've been ...	deltafaucet	Indianapolis, IN	4	1
33	Eleganza is a contemporary collection with min...	LacavaBathroom	Chicago, IL	3	0

Figure 1.2: Table of Data Collected

Sentiment Analysis:

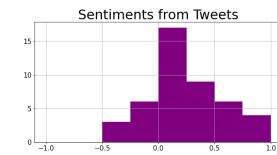


Figure 1.3: Distribution of Sentiment Score of Text from Tweets (Negative: -1.0 to Positive: 1.0)

Text Classification:

- Built a "Zero Shot" Classification Model using HuggingFace
- Used it to filter relevant tweets
- "Zero Shot" model is a
- Form of unsupervised learning
- Categorized tweets based on different topics

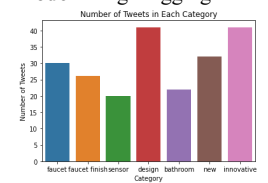


Figure 1.4: Tweets Per Topic

Image Classification:

- Built an image classification model on a test dataset using Keras and TensorFlow to serve as a building block for next year's students



Figure 1.5: Some Images from Tweets Collected