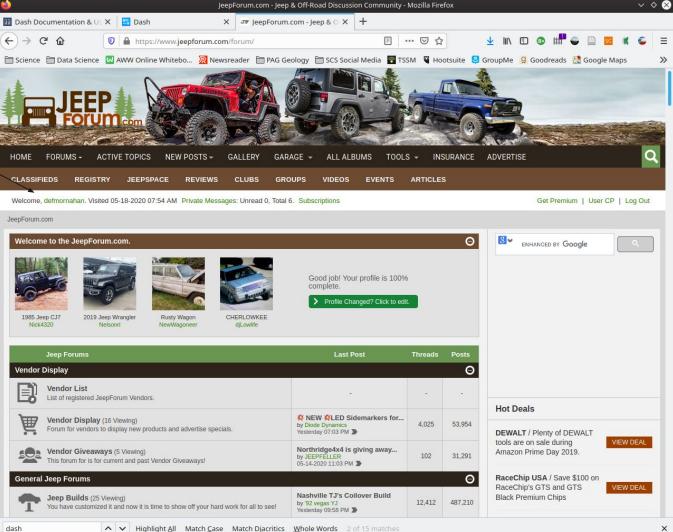
Paul Giesting Metis Data Science Project 4

Topic Modeling: jeepforum.com



Because I'm not just an analyst, I'm also a client. I mean, user.

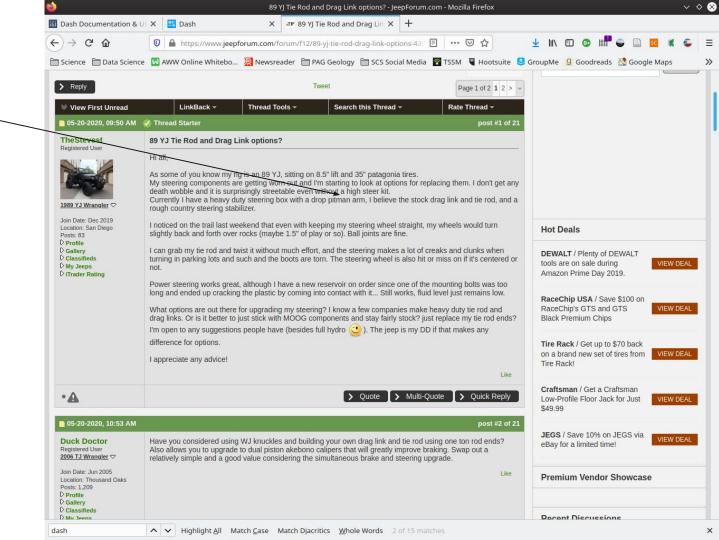




Big Picture on Topic Modeling:

- Understand topic breakdown for ad targeting.
- Monitor changes in topic breakdown over time.
- Recommendation system for users
 - ...especially if integrated with forum search.
 - Potentially a completely different way to navigate the site compared to the traditional reverse chronological discussion forum layout.

Note how much technical language there is here. Lots of opportunity to differentiate topics.



Toolchain







mongoDB.





- MongoDB Google Compute
 - Sampled ~47,000 records for analysis evenly spaced in time
- spaCy
 - POS
 - lemmatization
- Scikit-learn
 - Custom stopwords
 - Count vectorization
- Topic modeling
 - CorEx (unanchored)
- Plotly Dash (overlay Flask)

Further work

- Topic breakdown over time
- Topic breakdown by user
 - Recommender
- How to get probability values for existing CorEx topics on new documents?
- Extend app with time info, recommender
 - Monthly data update

Appendix

CorEx topic list

- Top 20 words per topic

- ['0: wire, fuel, engine, light, battery, sensor, check, plug, switch, ignition, wiring, gauge, idle, pump, relay, power, computer, fuse, harness, spark',
- '1: lift, spring, shackle, tire, suspension, leaf, inch, shock, soa, kit, stock, rear, ride, body, flex, bds, rc, perch, bl, leafs',
- '2: frame, hole, bolt, mount, door, rust, cut, metal, weld, drill, seat, paint, plate, piece, windshield, hold, tub, steel, spray, upper',
- '3: forum, jeepforum, https, php, showthread, thread, read, search, manual, write, information, haynes, showt, info, hayne, index, faq, google, chilton, magazine',
- '4: carb, intake, exhaust, air, manifold, run, filter, weber, system, cat, carter, nutter, emission, header, pipe, flow, motor, bypass, muffler, tank',
- '5: money, buy, spend, price, pay, people, trail, live, save, stuff, rig, love, street, deal, mud, daily, happy, plus, dd, decide',
- '6: start, oil, time, leak, work, change, clean, gas, water, right, let, cold, warm, gasket, week, today, notice, pan, blow, minute',
- '7: case, gear, shaft, transfer, tranny, slave, transmission, cylinder, shift, master, fluid, ax, bearing, swap, brake, trans, drop, housing, diff, slip',
- '8: axle, use, locker, end, dana, arm, pitman, big, set, ford, tie, far, cheap, cost, upgrade, probably, rod, different, option, carrier',
- '9: problem, turn, clutch, replace, pull, cause, fix, noise, pedal, happen, sound, stop, pressure, push, issue, engage, position, idea, bleed, play',

- '10: drive, help, speed, thank, road, mile, guy, appreciate, highway, vehicle, high, low, mph, feel, owner, greatly, mileage, wrangler, hill, ok',
- '11: jpg, mean, photobucket, learn, hot, albums, guess, heater, core, course, kinda, cool, note, crap, sorry, important, combustion, safety, word, prove',
- '12: remove, nut, tighten, screw, wrench, come, socket, tool, loose, break, hammer, loosen, inside, head, belt, torx, tight, heat, hand, tap',
- '13: sye, driveshaft, angle, pinion, cv, question, yoke, joint, shim, instal, point, install, vibration, ask, vibe, answer, driveline, correct, length, measure',
- '14: black, red, color, green, white, blue, yellow, radio, ebay, cgi, ebaymotors, tan, cb, stereo, interior, orange, boy, lifetime, pep, cloth',
- '15: line, vacuum, hose, cap, valve, distributor, rotor, connect, vac, radiator, pcv, firewall, vent, port, block, follow, dist, hei, egr, short',
- '16: ago, little, year, day, month, bit, long, hard, easy, hour, home, maybe, actually, winter, real, summer, area, rain, job, snow',
- '17: bar, flare, fender, track, wheel, sway, tj, roll, rim, bumper, rub, wide, fit, trim, steering, flat, speaker, style, trac, backspacing',
- '18: new, find, old, local, shop, store, yard, place, auto, pick, pretty, junk, dealer, replacement, original, repair, charge, purchase, free, model',
- '19: view, email, site, link, sell, pm, list, martin, build, pic, quadratec, winch, member, joe, shipping, ship, ft, page, protect, web']