

Jason Melbourne, Ph.D.

2535 Beverly Ave Apt C, Santa Monica, CA 90405
(831)-332-4607 jason.melbourne@gmail.com

TrueCar – Big Data Developer, Tech Lead, and Data Scientist (May 2013 – April 2016)

Tech Lead and Big Data Developer, *TrueCar Pricing Model Curves* (2015 – 2016)

The public face of TrueCar is the price curve, transaction histogram, and price trend shown for every vehicle on the site, e.g., <https://www.truecar.com/prices-new/toyota/camry-pricing/>. I was both the tech lead and one of the primary big data developers for the new localized pricing shown on the site today.

- Used Hadoop to store and enrich three plus years of vehicle transactions in Avro format.
- Used Java MapReduce to aggregate transactions across facets of time, space, and vehicle properties.
- Implemented a linear regression model with pluggable features to calculate the final price estimates.
- Published the pricing documents to Elasticsearch (ES) for consumption by APIs and analytics
- Built dashboards in Kibana to track transaction history, pricing model quality, and basic stats about the pricing documents
- Used the ES documents, Kibana dashboards, and Pig to identify data issues that were not surfaced through unit tests
- Used Oozie to manage the workflow of the project
- Organized the efforts of eight developers over the course of the project.
- Legacy SAS/SQL version took ~17 hours to complete and was run once a week, now completes ~30 min in hadoop, with another 40 min to publish the 1.5 million records to ES, and it runs once a day
- Provides truly local pricing for first time

Mentor and Data Scientist, *Automated Image Classification* (2015)

During the summer of 2015, I mentored a group of six brilliant summer interns. This team utilized TrueCar's massive vehicle image collection and implemented advanced computer vision algorithms, to answer the following questions: (1) where is the car, (2) is it the interior or exterior of the car, (3) what angle is the car positioned, (4) is there an advertisement on the image, and finally (5) what is the make and model of the car.

- Implemented OpenCV, GrabCut, to segment the image and pull out vehicle
- Implemented SIFT and SURF algorithms to extract basic image features, e.g., gradients and angles
- Used Bag-of-words framework with k-means clustering to group and ultimately classify images as exterior full, exterior detail, or interior shots.
- Used AWS GPUs and CNNs for specialized classifications
 - Start with Google trained image classification CNN
 - Retrain final layer of the CNN using large TrueCar imaging set
 - Classify position angle, accuracy ~90%
 - Classify make/ model ,accuracy ~85%, ~20% better than best published academic result, largely because of size of training set.
- Used Google Tesseract optical character recognition (OCR) and a library of words to determine if ads are on the images.

Big Data Developer, Hadoop Inventory Processor (2014)

The Hadoop Inventory Processor was the first big data project at TrueCar. We ingest ~10 million vehicle records from ~100 different data providers in Hadoop, normalize and enhance the data with Java MapReduce, and publish to Elasticsearch for consumption by APIs, including advanced search options. This process powers TrueCar's Used site, e.g.,

<https://www.truecar.com/used-cars-for-sale/listings/>. It runs in ~1 hour 6 times a day.

- Wrote the oozie workflow
- Wrote the Java MapReduce data enhancement job
- Used Pig and Hive to uncover data issues
- Previous SQL system ran in 8 hours once a day, Hadoop version runs in ~1 hour 6 times a day
- Previous system took about a week to add new provider + developer work; new system can add provider in ~ 1 hour and no developer work required (adding new providers now handled by the dealer data team).

Tech Lead, and Developer, Vehicle Recommendation Engine, Build data, and Automated Offers (2013 – 2014)

Once a customer registers with TrueCar they receive automated vin-based offers on vehicles that match their search preferences. Automated offers require two data sets, inventory to match on, and extremely accurate pricing for that inventory. I was the tech lead for automated offers and the primary dev for the vehicle recommendation engine that does the matches.

- Implemented a vehicle recommendation engine to match customers to dealer inventory
 - Weighted distance scoring algorithm based on vehicle features and customer preferences.
 - Implemented multiple “rankers” (e.g., weighting schemes) to test and tune for optimized close rate
- Lead the effort to incorporate Build Data into company
 - Build data is a vin-based record that details what the trim, color, options and price of a car is
 - Build Data is required for an producing an accurate automated offer
 - We purchased build data from 3rd parties
 - Set up vital systems for validating and enhancing the data, particularly price components where the vendor often had incomplete data
- Organized the team constructing the final offers
 - APIs construct the offers and store them in SQL
 - SQL stored procedures provided the final pricing based on info from build data, incentives, regional fees etc.

Developer and Data Scientist, TrueCar Mobile PriceCheck (2013)

Mobile PriceCheck is a feature on the TrueCar mobile app that allows customers to see the TrueCar price for a vehicle when they are on the dealer lot. I implemented the optical character recognition (OCR) routine that allowed the user to snap a quick photo of the vin number of the vehicle they were looking at and get the price.

- Wrote python API service for vin – OCR
- Used OpenCV to smooth and filter image pixels before running OCR
- Used Google Tesseract to perform OCR
- Led on-the-dealer-lot testing of the entire mobile PriceCheck experience including the OCR portion

Caltech – Research Scientist, Project Manager (2007 – 2013)

Principal Investigator, *Local Group Infrared Cluster Survey (LoGICS)*, Caltech (2011 – 12)

- Assembled team from 7 institutions to track stellar evolution in star clusters
- Authored 2 successful proposals for Adaptive Optics imaging from Keck Observatory
- Organized and led teams of 3 observers to obtain the data
- Mentored a team of students to analyze the data
- Presented results at the 2012 Keck Science Meeting

Research Scientist, *Bootes Research Team*, Caltech (2007 – 2012)

- Led 4 papers on the morphologies of infrared luminous galaxies
- Organized monthly group telecons
- Authored 4 successful proposals for Adaptive Optics imaging/spectroscopy from Keck
- Organized and led teams of up to 5 observers to obtain the data
- Presented results at international conferences in Madrid, Vienna, the USA, and Russia

Instrument Scientist, *NIRES Instrument Team*, Caltech (2010 – Present)

- Led testing of a near-infrared detector for a new spectrograph for Keck Observatory
- Determined a detector upgrade was necessary to achieve the instrument specifications
- Developing the User Interface for controlling the instrument at the telescope
- Writing a data reduction pipeline

Postdoctoral Fellow, Physics, California Institute of Technology (2007 – present)

- Organized funding and instructors, and led the Keck Adaptive Optics Workshop
- Initiated an effort to expand professional development of postdocs, exposing them to inquiry-based teaching techniques (see Grants below).
- Organized the poster presentations for the Keck Observatory's 20th Anniversary Gala

Special Projects

Spark Day (2015 – 2016)

- Organized 3 events to start blending our data science and data engineering teams onto a common platform of Scala and Spark. (1) Intro to Scala and data frames, (2) Intro to Scala Spark and SparkShell, and (3) Intro to Spark on AWS, with a hands on workshop

Hadoop Lunch (2014 – 2015)

- Led weekly discussions about the Hadoop Big Data architecture to develop a community of practitioners at TrueCar.
- Led hands on workshop on Java MapReduce attended by 20 data scientists and engineers
- Helped organize several hadoop hackathons around Hive, Pig, and MapReduce

arXivNow, released at www.mybigscience.com/arxivnow.html (2011 – 2013)

- I built an Objective-C application to quickly display and filter the daily astronomy literature
- arXivNow also provides a database to store and comment on interesting articles

Skills

Coding: Java MapReduce, Hadoop, Yarn, Oozie, Pig, Hive, HBase, Elasticsearch, Kibana, Python, Java APIs, IDL, Objective-C & Cocoa

Statistics: Linear Regression, Maximum Likelihood, Model Fitting, KS-tests, Correlation

Image Analysis and Computer Vision: OCR, Edge Detection, Fourier-Transforms, Filtering, Classification, Flatfielding, Flux Calibration, PSF Reconstruction, Convolution, CCDs, Infrared Detectors

Education

Doctor of Philosophy, Astrophysics, UC Santa Cruz (2006); **Master of Arts**, Astronomy, Wesleyan University (2001); **Bachelor of Arts**, Physics and Astronomy UC Berkeley (1995)