Hi there!

Congratulations, you have been shortlisted for interviewing further with SocialCops, and now have reached a Technical Round for our internship program!

As an intern at SocialCops, you will work on our Alternative Data team. Our Alternative Data Team builds solutions to unearth the ground truth about the real world, kilometre by kilometre - right from measuring affluence at 1 km grids around the world, to predict where malaria will break out next month.

From satellite data to government reports, from structured, internal data to unstructured, external data, and from online PDFs to paper surveys, the data sources we use are broad and varied. We integrate 200+ global data sources across different sectors - agriculture, demography, infrastructure and consumer affluence, for targeted, granular insights.

As an intern at SocialCops you would get to deal with diverse data, ranging from satellite data to sales data of big companies. You will be responsible for data modelling, cleaning, structuring, and handling large-scale datasets - working closely with our data scientists and economists to deliver holistic & scalable solutions. You will play a key role in converting a variety of messy datasets into clean and structured datasets by creating quality metadata files, running scalable R/Python scripts to model the data and perform data validations. You will also carry data analysis, create data visualizations, and create data models to make sense of data to power critical decisions.

As next steps, we are sharing one challenge statement. It is deliberately structured as an open-ended problem statement and is designed to help us understand your natural strengths & skills.

Challenge 1: Agriculture Commodities, Prices & Seasons

Aim: Your team is working on building a variety of insight packs to measure key trends in the Agriculture sector in India. You are presented with a data set around Agriculture and your aim is to understand trends in APMC (Agricultural produce market committee)/mandi price & quantity arrival data for different commodities in Maharashtra.

Objective:

1. Test and filter outliers.
2. Understand price fluctuations accounting the seasonal effect
    1. Detect seasonality type (multiplicative or additive) for each cluster of APMC and commodities
    2. De-seasonalise prices for each commodity and APMC according to the detected seasonality type
3. Compare prices in APMC/Mandi with MSP(Minimum Support Price)- raw and deseasonalised
4. Flag set of APMC/mandis and commodities with highest price fluctuation across different commodities in each relevant season, and year.

Data: https://drive.google.com/drive/u/0/folders/0B-zoMsiXW40gZlNtNnlINEszRTg

Variable description:

- msprice- Minimum Support Price
- arrivals_in_qtl- Quantity arrival in market (in quintal)
- min_price- Minimum price charged per quintal
- max_price- Maximum price charged per quintal
- modal_price- Mode (Average) price charged per quintal

Submissions**:**

1. Final cleaned file(s). (Bonus: if the files are shared using GitHub with well-versioned log)
2. Documentation around the methodology, analysis, and final results that you want to share with the Government of Maharashtra. Do use graphs and charts to substantiate your analysis. (Bonus- if you use GitHub pages / RPubs / etc. to share your documentation)
3. Script(s) and their documentation. (Bonus - using Jupyter Notebook or GitHub ReadMe.)
4. Visualisations, if any. (Bonus - if you use interactive dashboards)

Tip: The objective is to understand trends in each APMC and commodity separately so doing the analysis for just 1 commodity or APMC won't help.