

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 two challenge statements:

- (1) Challenge 1: Agriculture Commodities, Prices & Seasons
- (2) Challenge 2: Improving Healthcare in Maharashtra

From the 2 challenge statements, **pick one to solve**. These challenge statements are deliberately structured as open-ended problem statements and are designed to help us understand your natural strengths & skills. Pick the problem that you think you're best suited to solve & where you'll be able to send us the best output!

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-zoMsiXW40gZINtNnIINEszRTg>

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)

Challenge 2: Improving Healthcare in Maharashtra

Aim: Your team has begun an exploratory project to map the healthcare scenario around the world - starting with Maharashtra. You are presented with a data set to identify the root causes of the extreme health trends in the Nagpur district. Your aim is to share this analysis with policymakers, private hospital chains & NGOs to improve healthcare in the district.

Objective

You are tasked with building the Analytics engine using the data recorded under the flagship program: National Rural Health Mission (NRHM). A sample of the data for NRHM Maharashtra, Public Healthcare Centre (PHC), has been shared (Excel). Data Download Link: <https://drive.google.com/drive/folders/0B0-TJY1UhiGpWINLWEIwOEpHSFU>

The policymakers are keen to understand the possible reasons associated with these health trends, in order to realign scheme implementation efforts to the quality of services vs. extent of coverage.

Background

As an intern on the team, you are requested to do the following:

1. Clean, structure, and prepare the data for the analysis so that you can compare all the files together.
2. In the decentralized healthcare scenario, Sub-centre makes the lowest division, followed by PHC and then block level hospitals (or RH). Tell us if the numbers follow this vertical hierarchy.
3. Report 10 important health trends and visualize them using different trend graphs. Brownie points for Causal analysis!

Submissions:

1. Final cleaned file(s). (Bonus - if the files are shared using GitHub with well-versioned log)
2. Indicators that you are using as an Excel Sheet.
3. Word doc around the trends, and final results that you want to share with the Government of Maharashtra. Use visualisations to show analysis and your final results. (Bonus- if you use GitHub pages / Rpubs / etc. to share your documentation and share interactive visualization)
4. Script(s) and their documentation (Bonus - using Jupyter Notebook or GitHub ReadMe.)

Notes:

- An internal document on scheme can be found here: [NRHM NRHM](#)
- You will get an overview of the central scheme here: <http://nhm.gov.in/nhm/nrhm.html>

- It's crucial to understand the decentralized healthcare framework in India (Sub-Centre, PHC, District Hospital, Rural Hospital, Referral hospitals), and the care-givers (ASHA, ANM) involved.
- The focus of the scheme is documented here:
<http://nhm.gov.in/nhm/about-nhm/goals.html>
- Follow this link to get an idea of what of the Maternal Health status and care-giving framework: <http://nhm.gov.in/nrhm-components/rmnch-a/child-health-immunization.html>
- Follow this link to get an idea of what of the Maternal Health status and care-giving framework is:
<http://nhm.gov.in/nrhm-components/rmnch-a/maternal-health/background.html>
- The following links will give you an idea about the current reports generated for monitoring purposes by the Ministry: (1)
https://nrhm-mis.nic.in/hmisreports/fmstandard_reports.aspx ; (2)
<https://nrhm-mis.nic.in/hmisreports/analyticalreports.aspx> (3) ;
<https://nrhm-mis.nic.in/Pages/HMIS-PeriodicReport.aspx>
- Ensure that you understand what are important health indicators globally - read up on UN and World Bank reports.
- The task focus is to answer questions such as: What are possible causes of an increasing Maternal Mortality Rate in Maharashtra? Is there an impact of Anemia on Maternal Mortality? Does increasing Immunization coverage during Antenatal Care (ANC) help in reducing # maternal deaths during/post pregnancy?
- Cover recent news articles to find recent metrics that were quoted and their trends - this will give you an idea on what is crucial to assess the scheme's progress.
- The Ministry officials are very well versed with NRHM implementation. Your suggestions should be fine tuned to technicalities of scheme implementation.