Seattle Women in Data Science (WIDS)
Interview Preparation Group

Table of Contents

1. Background

This group is aimed at supporting individuals in the preparation for data science and data analyst interviews and more importantly to help people get the data job they want. Given the expansive nature of tasks, knowledge and competencies that can fall into data science and data analyst jobs, and the degree to which the field still remains undefined, preparing for these interviews can be challenging. The objective is to provide an opportunity to interface with others to improve comfort with key concepts and approaches to interviews that can help achieve success in data science and data analyst interviews.

People joining the group should be relatively confident in their level of preparedness for interviewing within the next four months based on their data experience and expertise. This group is not aiming to target people that are new to data analytics, data science jobs and just are beginning to consider entry into the field.

Note: This group is specifically focused on people that are aiming for analytical and quantitative tracks of data science. Interview preparation for tracks that are more focused solely on data engineering and advanced algorithm design should refer to Puppy and She's-Coding meet-up groups.

2. Slack Channel: Women in Data Science

General announcements for this group will be made through the following WIDS sub-channels

a. #interview-prep

This channel is designed to provide a community in which to share and discuss important topics about data science interviews.

b. #interview-prep-org

This channel is designed to provide a community in which to discuss new and interesting project ideas that can help achieve the mission of the group to better prepare people for data science interviews and to discuss important organizational objectives such as engagement of volunteers.

3. Target Audience

- Individuals preparing for data science and data analyst interviews within the next four months
- Individuals who want to maintain competencies in broad key concepts typically covered in data science interviews

4. Data Science/Analyst Tracks

- Statistics
  - Statistical Distributions (poisson, logistic, gaussian/normal, bernoulli)
  - Hypothesis Testing
- Machine learning/statistical learning
  - Parametric
    - Linear/General Additive Models
    - Logistic
    - Lasso and ridge regressions
  - Non-parametric
    - SVM
    - Decisions Trees and Random Forest
    - Gradient boosting methods
    - Clustering (K-means and others)
  - Neural networks
    - Feedforward Neural Network
    - Radial basis function Neural Network
    - Kohonen Self Organizing Neural Network
    - Recurrent Neural Network (RNN)
    - Convolutional Neural Network (CNN)
    - Modular Neural Network
  - Natural Language Processing (NLP)
- A/B Testing
- SQL
  - Statistic computations
- Basic analytical algorithms
  - monte-carlo simulation
  - regression using for loops

5. Logistics and Group Structure

To get involved please read-over this document, check out our [github repositories](#) and fill out the [survey form](#)

a. Short-term

Sign-ups and participants to the group will occur on an on-going basis. However, because of limited volunteer capacity the available slots will be limited to people that are actively preparing for data science interviews and are able to make valuable contribution to the group dynamics.

- Meetup nights for selected set of participants
    - Hacknights for portfolio development and contribution to group
    - Practicing interview questions within 4 key areas
    - Resume and portfolio review advice
    - Blog development to get group and you noticed
- Meetup nights for everyone that may be more general discussions or forums for interview preparation

b. Long-term

The long-term goal is to have a group that is inclusive to the larger population of people that are aiming to interview for data roles. Matches will try to be automated with people being assigned to other group participants based on their current level of skills and expertise and their data preparation interests.

The matches will be expected to be informal and held via group video-chats or in-person. The main objective of the organizer(s) are to
- Provide resources and structure for practicing for data science interviews
- Match individuals to people with similar goals and strengths
- Supplementing groups with additional complementary help based on demand/availability
- Tracking activity and re-organizing groups based on fluctuations in participation activity.

The informal group meetings may be supplemented on occasion with more formal seminars or support.

6. Participant Matching

Matches will be developed based on backgrounds of interested parties. Using response to key questions and/or participants CVs matches will be developed to sort people into different groups where ultimate job/learning goals are aligned. It will be an aim to try to supplement interview practice with people in areas that a participant is looking to improve upon with people from the pool of participants and volunteers who have specialized expertise (e.g. coding, statistics, neural networks etc.).

The following are broad criteria that will be used in develop participant matches. Of course, the quality of matches may be differ depending on the pool of people that are participating in the group at any given time. People interested in participating should fill out the following participant survey. Key criteria used in matching will be:

- Interview stage
    - Review/Preparation
    - Actively interviewing
- Target positions
    - Small, medium, large
    - Job ad types/key tasks
        - quantitative
        - image processing
        - natural language processing
- Educational background
- Experience
- Primary code specialization
    - SQL
    - R
    - Python
- Occupational targets
    - quantitative inference
    - machine learning
    - data engineering
- Availability
- Specializations
    - quantitative
    - image processing
    - natural language processing
- Degree of competency over different areas (statistics, general machine learning algorithms, SQL)

- Degree to which person is interested in improving in areas (statistics, general machine learning algorithms, SQL)

7. Volunteering

a. Proposed Plans for Engaging Volunteers

❏ Looking for volunteers via hackathons such as [Seattle-Democracy-Lab](#) or https://www.volunteermatch.org/
❏ Reach out to bigger meet-ups and have the organizers mention about the group (e.g. Puppy, Seattle Deep Learning Meetup, and/or She-Coding, Seattle Girl Geek Dinner). On the same lines, have the brand with you (Company sponsored meet-ups). That will get us more volunteers.
❏ Tie up with interview preparation sites like [pramp](#) and [interview.io](#). In case it is a paid one, we could reach out to them for discounts/few free interviews prep calls.
❏ Search on linkedin for DS aspirants and approach them.
❏ Develop a FB page for meetup/group which easier for people to reach out
❏ Blog about the data science interview process and projects in a Medium article.
❏ Catch hold of influencers. Have them write an article, share in bigger meet-ups or tweet about the support group.
❏ Talk to DS bootcamps in Seattle know about the support group.
❏ Expand support group to larger population set (women and men).
❏ Develop expertise to guide people from our own experience and identify people working as a Data Scientist that can be expert references for the support group.
❏ Get people from the Machine learning study group and bootcamp graduates to migrate to this group eventually.
❏ Come up with a list of data science interview preparation articles.
❏ Potentially mandate some small inputs/tasks (e.g. finding X number of new data science questions that can help people in preparation as minimum requirement for participation in group)
❏ Placing projects on github that allow volunteers to showcase their contribution
❏ Having more extensive hack sessions (e.g. full-day on weekend) or alternatively joining one of the bi-monthly hackathons through
❏ Gathering data on different articles on the data science interview process -- can be rolled out into more detailed project (recommendation system).
❏ Automating process of volunteering

b. Volunteer Needs

Volunteers are important part of ensuring this group is able to work and be maintained. Immediate needs include the following:

- ❏ Development and expansion of Data Science Interview Database
  - ❏ This could involve identifying website that are feasible for usage and scraping of databases and employing an algorithm for question classification.
- ❏ Development of Database of Reference Documents that can advise people on the best resources to use for preparing for certain areas of data science interviews.
- ❏ Comments or Feedback on
  - ❏ Match Criteria
  - ❏ Supplemental Resources
  - ❏ Additional support
- ❏ Generation of an algorithm for matching people together
- ❏ Ideas for helping to easily track and keep on top of participants progress in the interview search and practicing to better ensure goals are being met
- ❏ Interview coaches and people that can provide expertise for people in certain areas of interview preparation
  - ❏ Statistics
  - ❏ A|B Testing
  - ❏ Coding
- ❏ Github project/portfolio development

8. Reference Documents

This section provides some key resources that can be helpful in the preparation for Data Science Interviews.

- ● Data science questions database
  - ○ Starting point: https://github.com/kojino/120-Data-Science-Interview-Questions/blob/master/predictive-modeling.md
  - ○ Database of questions falling into different types (work-in-progress).
    - ■ Probability
    - ■ Hypothesis testing
    - ■ Case study
    - ■ Machine learning algorithms
    - ■ Experience questions
    - ■ Situational questions

- Each month groups to highlight projects in informal setting.
  -
- Books
  - [Introduction to statistical learning](#)
  - [Elements of statistical learning](#)

9. Contacts

a. General

seattle.data.interview.prep@gmail.com

b. Primary

Natalie Chun (natalie.l.chun@gmail.com)
Vandana Iyer (vandy.iyer@gmail.com)