CptS/CS/STAT 424: Data Analytics Capstone

Syllabus

Course Information
Course name: CptS/CS/STAT 424: Data Analytics Capstone
Credit hours: 3
Semester: Spring 2019
Prerequisites: Senior standing in Data Analytics major
Meeting times and location: TuTh 1:25—2:40 pm, EME B46
Course website: https://scads.eecs.wsu.edu/index.php/data-analytics-capstone/

This public course website will be used to post the syllabus. Additionally, the course management platform OSBLE+ (https://plus.osble.org) will be used for posting course materials, assignments, announcements and messages; and for handling assignment submissions and instructor feedback. The OSBLE+ course site will be accessible only to students enrolled in the course.

Instructor Information
Instructor: Assefaw Gebremedhin
Office: EME B43
Email: assefaw DOT gebremedhin AT wsu DOT edu
Homepage: https://www.eecs.wsu.edu/~assefaw/
Office Hours: Wednesdays 12—1:30 pm, or by appointment.

Course Overview and Course Goals
The Capstone course aims at providing students with an opportunity to integrate and apply the algorithms, methods and tools they have learned throughout the program to solve real-world data analysis problems that have an interdisciplinary nature. Students will conduct a team-based project that involves the main aspects of the data analytics process, and will submit a consolidated report and give a presentation at the conclusion of the project. The course serves as a final preparation for students entering into the profession. Students get experience in working as teams, participating in project planning and scheduling, writing reports, giving presentations, and interpreting results in a professional manner.

Project
A capstone project is sponsored by a company or by an academic research group. A typical project is expected to map to the following overall flow (the exact nature of data analysis and output will depend on the questions and objectives, as well as the application domain):
1) Engage client to establish questions and objectives;
2) Critically review technical literature relevant for the project;
3) Organize, clean and pre-process data;
4) Exploratory Data Analysis;
5) Data Analysis-based Inference/Prediction/Explanation of the studied phenomenon and/or Recommendation of decision/action based on analyzed data;
6) Data summarization and visualization; data product preparation;
7) Communicate key insights to a non-technical audience with the final data product, including the written report and oral-digital presentation.
**Student Learning Outcomes**

Through successful completion of their capstone project, students will be able to:

1. Identify relevant questions and objectives through client engagement;
2. Demonstrate information literacy through a critical review of technical literature relevant for the management and analysis of data for their group project;
3. Develop a project-appropriate plan and structure for data management;
4. Resolve group work allocation, leadership and cooperation issues;
5. Structure, manage and access one or more large, complex datasets;
6. Complete the analysis and interpretation of a complex, real-world data project; and
7. Present the analysis and interpretation of a complex, real-world data project in both written reports and digital+oral presentations.

**Table 1. Assessment of Student Learning Outcomes**

<table>
<thead>
<tr>
<th>SLO</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Writing Assignment #1, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>2</td>
<td>Writing Assignments #1, #2, #4, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>3</td>
<td>Writing Assignment #2, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>4</td>
<td>Peer Grading, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>5</td>
<td>Writing Assignment #3, Presentation # 1, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>6</td>
<td>Writing Assignment #4, weekly instructor progress grade, mentor evaluation</td>
</tr>
<tr>
<td>7</td>
<td>Writing Assignments, Presentations</td>
</tr>
</tbody>
</table>

**WSU Student Learning and UCORE Goals**

- **Depth, Breadth and Integration of Learning**
  Students will demonstrate depth, breath and integration of learning by applying the data analytics concepts and techniques learned throughout the academic program to a “real” academic or professional scenario or problem related to their application area or track.

- **Critical and Creative Thinking**
  Students will demonstrate critical and creative thinking by formulating questions, hypotheses and objectives through client interviews; determining data collection and/or compilation needs for their specific project; creating and implementing a project-specific data processing and management plan; conducting an exploratory data analysis; analyzing their data using a range of techniques learned in the program; and interpreting their results.

- **Communication**
  Students will be required to write four reports, one of which is a major project synthesis, and make two professional presentations.

- **Information Literacy**
  Students will demonstrate ability to review relevant technical literature for a data analytics project and appropriately cite sources. Examples of technical literature include peer-reviewed journal articles, white papers, tutorials, standards, application notes, and project related documentation provided by client of the project or mentor.
Additional learning goals appropriate to the discipline
Student progress toward the academic program learning objectives (below) will be evaluated using a student self-assessment tool implemented near the end of the course.

1. Attain synthetic knowledge of Big Data analysis from computer science, statistics, and an application field.
2. Demonstrate the technical skill to customize existing data analytics software packages for the management, processing and analysis of very large, complex data sets.
3. Demonstrate the analytical skills to appropriately apply statistical and data mining techniques to explore, summarize and make inferences from large complex datasets.
4. Demonstrate critical thinking skills needed to appropriately interpret data analysis problems, formulate questions, and interpret the results of big data analyses.
5. Effectively communicate data analytics techniques and results to both the academic community, professionals in one application field, and the general public in writing, discussion, digital media, and other communication formats.
6. Demonstrate an understanding of the professional, ethical, legal, security, and social issues and responsibilities associated with the field of data analytics.

Teams
Teams will initially consist of 3 to 5 students and a team will receive their primary project-specific guidance from a mentor provided by the sponsoring company or academic research group. It is important that a student is placed onto a project that will utilize their unique data analytics skills and be relevant to their track. Similarly, it is important that students bring and deploy data analytics skills and field-specific knowledge to the team for achieving the project objectives.

Evaluation of Student Work: Both team process and subsequent products will be evaluated. Process and products will be evaluated based on the team's abilities to:
1. Identify the questions to be answered and/or data products to be developed;
2. Develop and implement a data management system for the project;
3. Clearly articulate appropriate analytical methods for the project questions and products;
4. Implement the appropriate analyses and interpret the results;
5. Communicate in an organized and professional manner with multiple audiences.

Yellow Slip: Students must demonstrate competence at 1) functioning within teams and 2) understanding professional and ethical responsibilities. The yellow slip is a way for team members to draw the instructor’s attention to a member who exhibits behavior that contradicts this competence. Until semester grades are submitted to the WSU Registrar, students can issue a “yellow slip” to a team member and email it to the instructor. The result of a yellow slip can be severe enough to change the letter grade.

Student Work Load
CptS/CS/STAT 424 is a 3-credit course. The 3-credit designation normally implies that on average the student is expected to spend 3hrs ("lectures") + 6hrs ("homework") = 9 hours per week working on the course.

The Tuesday and Thursday 1:25—2:40pm class meeting times will be used for a few lectures in the beginning (and later in the semester as needed) but will otherwise be used for discussing project progress. The 6 hours per week outside of class and team meetings with the mentor will be spent on
the following activities: 1) additional team meetings; 2) completing writing/presentation assignments; 3) reading the technical literature related to the project; 4) meeting with clients as arranged by the mentor; 5) helping with project management; 6) engaging in the analytical process; 7) organizing team headquarters including obtaining pertinent data and software; and 8) developing and obtaining approval for the project’s data analysis plan.

There will be four writing assignments and two presentation assignments in this capstone course. Please see the Appendix for a complete list of assignments.

**Capstone Final Presentation:** Each project team shall prepare a digital presentation for the team project and present it during the scheduled time for the course final exam. A team of judges from academic and industry clients will evaluate all Data Analytics Capstone presentations. Detailed guidelines on the presentation preparation will be made available on the course’s OSBLE+ website.

Project presentation will mainly be evaluated based on quality, content and design. A percentage of the presentation grade will be based on the teams’ performance during the presentation and associated question and answer period. The instructor will evaluate and grade the presentation. Scores from the clientele judges will NOT be directly factored into the course grade.

**Weekly Review Meetings**

Each team will meet once a week, typically with their mentor, and all team members will report their progress for the past week and present a plan for the upcoming week. A team may request participation of the instructor in a meeting. Each team member will prepare 1 or 2 slides summarizing their progress and plan. The team liaisons should make sure there is a link to the weekly presentation slides from the teams’ GitHub repo. A summary of the weekly meeting progress report will be presented to the entire class.

The instructor will evaluate and grade the progress within the past week based on the weekly meeting presentations and the activity at GitHub. If a team member cannot attend due to a valid excuse, he/she should notify the instructor and the team members. Skipping a meeting without notification will result in a score of “0” for the weekly evaluation. Students with more than three unexcused absences in weekly meetings will receive a course grade of “F”. The most successful teams schedule face-to-face team meetings at least on a weekly basis and they insist on 100% team attendance at these meetings.

**Repositories and Course Management Platform**

The project teams will primarily use GitHub for software version control, task management, bug tracking, and posting the meeting notes. Each team will have a private repository at the EECS GitLab server. The project teams should utilize GitHub tools and features as effectively as possible. The instructor and the industry mentor will monitor a team’s progress through its repository and they will consider GitHub activity in evaluating the teams’ progress during the semester.

The instructor will communicate with the class and project groups primarily via the learning management platform OSBLE+ ([https://plus.osble.org](https://plus.osble.org)). Students will create an account on OSBLE+ to be able to use the course’s site on the platform. The course site will also be used for posting course material and resources, assignments, announcements, and messages. In addition, the site will be used for handling assignment submissions and instructor feedbacks; that is, teams will submit their writing assignments on OSBLE+ (one submission per team is sufficient) and instructor feedback and assignment grades will be communicated through the platform.
Grading

All course material will be graded based on a scale from 1 to 5. “Score 5” means your work satisfies all the expectations. Above 5 is to recognize exceptional work. Extra credit will be given up to 0.5 points. All graded items (including writing assignments, presentations, data projects, meeting reports) will be evaluated using rubrics, found in the syllabus Appendix. Late assignments, unless prior instructor approval is obtained, will be penalized at a rate of 0.1 points/day.

Weighting of Graded Components:

1. **Writing Assignment #1**: Questions, Objectives & Review .......................... 10% (Team Grade)
2. **Writing Assignment #2**: Data Management ........................................... 10% (Team Grade)
3. **Writing Assignment #3**: Exploratory analysis ........................................... 10% (Team Grade)
4. **Presentation #1**: Data Analysis Plan ..................................................... 10% (Team Grade)
5. **Writing Assignment #4**: Final Report .................................................... 20% (Team Grade)
6. **Presentation #2**: Multimedia presentation ............................................ 15% (Team Grade)
7. **Weekly Progress Evaluations by the Instructor** ...................................... 15% (Individual Grade)
8. **Peer Evaluations** .................................................................................. 10% (Individual Grade)

100%

Peer Grading: At the end of the semester, each student will fill-in and submit a “peer evaluation” form. Students will be asked to provide feedback about the contributions of their teammates to the project and to the assignments, with a score of 100 meaning the student contributed their fair share and 0 meaning they did not contribute at all. The assignment scores (above) will be adjusted according to peer evaluations, i.e., adjusted assignment grade = team assignment grade (points × fraction of expected contribution (e.g. 80/100)). The instructor reserves the right to adjust the peer evaluation scores if the instructor and the mentor agree that the scores were prejudiced (either too high or too low).

Letter Grades:

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>A</th>
<th>A-</th>
<th>B+</th>
<th>B</th>
<th>B-</th>
<th>C+</th>
<th>C</th>
<th>C-</th>
<th>D+</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Score</td>
<td>93%</td>
<td>90%</td>
<td>86%</td>
<td>83%</td>
<td>80%</td>
<td>76%</td>
<td>73%</td>
<td>70%</td>
<td>66%</td>
<td>60%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>&lt;93%</td>
<td>&lt;96%</td>
<td>&lt;86%</td>
<td>&lt;83%</td>
<td>&lt;80%</td>
<td>&lt;76%</td>
<td>&lt;73%</td>
<td>&lt;70%</td>
<td>&lt;66%</td>
<td>&lt;60%</td>
</tr>
</tbody>
</table>

Protecting Intellectual Property (IP) and Confidential Data

Teams have an obligation to protect both IP and data that clients share with them. WSU employees, including faculty, staff and graduate students are legally bound to protect IP and confidential data. Do not post IP or data at non-password-protected websites. Questions about IP should be directed to the mentor or directed to WSU attorneys trained in IP issues. Similar comments apply to information that government and military entities label as “sensitive” or “classified”. Ask the instructor if you need contact information for WSU professionals working with these issues.
Table 2. Tentative Weekly Schedule

Below is a tentative schedule of activities in the course and topics of discussions. This will be adjusted as needed during the course of the semester. Some of the weeks may feature guest lectures by professionals from industry or academia.

<table>
<thead>
<tr>
<th>Week</th>
<th>Discussion</th>
<th>Assignments due by week end</th>
</tr>
</thead>
</table>
| 1    | • Course Introduction  
      • Mentors present project summaries | • Pre-course survey out |
| 2    | • Group formation & discussion  
      • Professional leadership and effective team work | • Survey in  
      • Groups formed |
| 3    | • Group progress reports & discussion  
      • Confidentiality & IP issues | |
| 4    | • Group progress reports & discussion  
      • Capstone Data Management Ethics | • Writing Assignment #1 due: Questions, Objectives & Review |
| 5    | • Group progress reports & discussion  
      • Feedback on Writing Assignment #1 | |
| 6    | • Group progress reports & discussion  
      • Exploratory Data Analysis | • Writing Assignment #2 due: Data Management |
| 7    | • Group progress reports & discussion  
      • Feedback on Writing Assignment #2 | |
| 8    | • Group progress reports & discussion  
      • How to construct and deliver quality presentations | • Writing Assignment #3: Exploratory Data Analysis |
| 9    | • Group progress reports & discussion  
      • Feedback on Writing Assignment #3 | |
| 10   | • Group presentations | • Presentation #1: Data Analysis Plan |
| 11   | • Group progress reports & discussion  
      • Critique of presentations | |
| 12   | • Group progress reports & discussion | |
| 13   | • Group progress reports & discussion | |
| 14   | • Group progress reports & discussion | • Writing Assignment #4: Final Report |
| 15   | • Group progress reports & discussion | |
| Finals | | Presentation #2: Multimedia presentation (Final Exam scheduled time) |
References

There is no required textbook for this capstone course. The following books are recommended as references for some of the topics:


- *R for Data Science* by Hadley Wickham and Garrett Grolemund. O'Reilly, 2017. ISBN 978-1491910399. The book is freely available online at: [https://r4ds.had.co.nz/](https://r4ds.had.co.nz/)

Students with Disabilities

Reasonable accommodations are available for students with a documented disability. If you have a disability and need accommodations to fully participate in this class, please either visit or call the Access Center (Washington Building 217; 509-335-3417) to schedule an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center. For more information contact a Disability Specialist on your home campus:

Pullman or WSU Online: 509-335-3417
[http://accesscenter.wsu.edu, Access.Center@wsu.edu](http://accesscenter.wsu.edu, Access.Center@wsu.edu)

Academic Integrity Policy

Academic integrity is at the heart of all higher education philosophies. Adhering to academic integrity policies ensures that you provide yourself with the best education possible. Maintaining academic integrity assures you receive the credit you deserve for your ideas.

You are expected to know and understand Washington State University Academic Integrity Policies. Copying and plagiarism of other sources will result in an automatic 0 or F on the assignment. For a second offense, an automatic F in the course will be awarded. University authorities will be notified and the proper procedures will be followed. Possible outcomes resulting from violating these policies include denial of certification into your program and expulsion.

Please thoroughly review the following website for more information about the WSU academic integrity policy: [http://academicintegrity.wsu.edu/](http://academicintegrity.wsu.edu/)

Please also review the following website for more information about the EECS academic integrity policy (if any discrepancies are found within the EECS and WSU academic integrity policies, always follow the WSU policy): [http://www.eecs.wsu.edu/~schneidj/Misc/academic-integrity.html](http://www.eecs.wsu.edu/~schneidj/Misc/academic-integrity.html)

WSU's AWARE Network

As a student you have many responsibilities and obligations. One of the most important obligations that you may have is to your friends and peers at WSU. If you feel like one of your friends or peers is struggling with academics because of physical or mental health please inform the instructor and/or appropriate university personnel. For more information refer to [http://aware.wsu.edu/](http://aware.wsu.edu/).
WSU's Campus Safety Plan

Campus safety is a priority for WSU. Please become familiar with the campus safety plan at: http://safetyplan.wsu.edu/. Also, review the Office of Emergency Management website at: http://oem.wsu.edu/. These sites will provide you with information about policies and procedures for campus safety. You should sign up for WSU alerts through http://alert.wsu.edu/

WSU Classroom Safety

Classroom and campus safety are of paramount importance at Washington State University, and are the shared responsibility of the entire campus population. WSU urges students to follow the “Alert, Assess, Act” protocol for all types of emergencies and the “Run, Hide, Fight” response for an active shooter incident (https://oem.wsu.edu/emergency-procedures/active-shooter/). Remain ALERT (through direct observation or emergency notification), ASSESS your specific situation, and ACT in most appropriate way to assure your own safety (and the safety of others if you are able).

Please sign up for emergency alerts on your account at MyWSU. For more information on this subject, campus safety and related topics, please view the FBI’s Run, Hide, Fight video (https://www.fbi.gov/services/cirg) and visit the WSU safety portal (https://oem.wsu.edu/about-us/).

Important Dates and Deadlines

Students are encouraged to refer to the academic calendar often to be aware of critical deadlines throughout the semester. The academic calendar can be found at http://registrar.wsu.edu/academic-calendar.

Weather Policy

For emergency weather closure policy, consult: http://alert.wsu.edu.

Changes

This syllabus is subject to change. Updates will be posted on the course’s website.
Appendix: Sample CptS/CS/STAT 424 Assignments for a typical Spring semester

<table>
<thead>
<tr>
<th>Assignment Generic Name</th>
<th>Assignment Descriptor</th>
<th>Average length</th>
<th>Terse summary of the prompt given to students for these assignments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Assignment 1</td>
<td>Questions, Objectives, &amp; Literature review</td>
<td>5 pages + appendices and images as needed</td>
<td>Work with data provider to clearly define questions and objectives for the data analysis. How does the data collected address these questions and objectives? Document the scope, approach, and schedule of the data analytics project. Critically review and cite key technical literature for project.</td>
</tr>
<tr>
<td>Writing Assignment 2</td>
<td>Data Management</td>
<td>5 pages + appendices and images as needed</td>
<td>Write a data management plan for this project, including a database design. Discuss how this plan will facilitate the needed analysis. Cite relevant literature.</td>
</tr>
<tr>
<td>Writing Assignment 3</td>
<td>Exploratory analysis</td>
<td>6 pages + appendices and images as needed</td>
<td>Report findings from exploratory data analysis. Cite relevant literature.</td>
</tr>
<tr>
<td>Presentation Assignment 1</td>
<td>Data Analysis Plan</td>
<td>20 minutes + slides &amp; digital data representation as needed</td>
<td>After presenting a short overview of the first three assignments, project teams will outline their data analysis plan, including methods, timeline, responsibilities and expected results. Cite key literature. Instructors will evaluate.</td>
</tr>
<tr>
<td>Writing Assignment 4</td>
<td>Final Report*</td>
<td>10 pages + appendices and images as needed</td>
<td>This written report covers all team activities during the class, including the data analysis, literature review, data projects and interpretation of results. This will serve as the UCORE assessment artifact.</td>
</tr>
<tr>
<td>Presentation Assignment 2</td>
<td>Multi-media Results</td>
<td>20 minutes + slides &amp; digital data representation as needed</td>
<td>Project teams will present findings. A team of evaluators from industry and academia will evaluate all Data Analytics Capstone presentations.</td>
</tr>
</tbody>
</table>

Detailed assignment descriptions and rubrics are listed on next page.

*Mentor Grading:* Based upon weekly progress slides and group meetings, the group mentor will assign each group member a grade of 0-10. This grade will be split equally between: (a) teamwork, conflict resolution, and effective communication; and (b) completion of individual responsibilities representing a fair share of group responsibilities. As with assignment grades, a score of 5.0 on either half of the mentor grade represents complete satisfaction, with up to 0.5 extra credit points possible for exceptional teamwork or assumption of more than a fair share of individual responsibility.
Assignment Descriptions and Rubrics

Below is a draft description in “generic form” of the four planned Writing Assignments and two Presentation Assignments in the course. When specific projects to be done in the semester are finalized, the assignments will be customized as needed. The final text of the assignments will be posted on the OSBLE+ site of the course.

Writing Assignment #1   Plan of research and data collection

Summarize your client’s data analysis problems, questions and objectives. What data products or answers does your client expect to result from your work?

Write a short critical review of the relevant technical literature for your project. This review will be revised and expanded over the course of the semester.

Write out a plan for how you would go about addressing your problem of interest. If it is an estimation issue, discuss what you want to estimate and why it is of interest. If it a test you want to conduct, discuss the hypothesis you want to test. Now, discuss a plan for collecting data for this problem. This data can be a result of an experiment, an observational study, or secondary data that you compile from various sources. This data has to be collected and or compiled by you from multiple sources. (A complete, processed dataset will normally not suffice for Capstone projects.)

Discuss your plan for collecting this data. What variables are you studying, and what sort of errors will you be facing in the data collection process? Talk about bias and other data related issues. Write a project report about 3-4 pages in length discussing the data collection, analysis and conclusions. Include a discussion section about things you would change. Write in the style of a technical paper. Include title, author names, abstract and the body of the paper.

This will be a group report, with individual contributions evaluated via anonymous peer feedback.

- **Rubric**

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment of problem, questions, objectives, data products and answers</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Critical literature review</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Evaluation of data needed to address above</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Data Collection plan</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Quality of writing and report organization</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>
Writing Assignment #2  Data pre-processing, transformation and management

Address the following towards preparing a clean dataset suitable for analysis and modeling.

• Data Cleaning
  - Identify missing values and develop suitable strategies for dealing with the missing values (e.g. strategies for deleting, imputing, using maximum likelihood);
  - Identify outliers and develop strategies for dealing with outliers (e.g. trim, include);
  - Determine if your numerical data needs special coding (e.g. binning, rescaling);
  - Determine if your character data needs conversion (e.g. to lowercase) and regular expression manipulation;
  - Ensure integrity of your data (entity integrity, referential integrity, type checking, format, bounds);

• Data Reshaping
  - Consider if your dataset behooves data reshaping maneuvers such as “subseting”, aggregating, sorting, and normalization. Identify effective mechanisms for achieving these.
  - Determine whether your dataset needs to be assembled from different sources and identify effective data frame mechanisms (e.g. set, join, and merge operations).

• Data Management
  - What metadata are available and needed? What are the relevant metadata standards?
  - How will data be structured for your analyses?
  - What are your policies for access, sharing and re-use?
  - Develop a plan for short-term management addressing version control, backups, security and responsibility.

• Cite relevant literature.

This will be a group report, with individual contributions evaluated via anonymous peer feedback.

• Rubric

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Cleaning</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Data Reshaping</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Data Management</td>
<td>1.9-2.0</td>
<td>1.6-1.8</td>
<td>0.0-1.5</td>
</tr>
<tr>
<td>Quality of writing and report organization</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>
Writing Assignment #3 Exploratory Data Analysis

For your “clean” and organized dataset address the following:

- Plot distribution of key variables;
- Create metrics/measurements/statistics that summarize the data;
- Find out if there are outliers/anomalies in your dataset, and if there are outliers, develop strategies for dealing with them;
- Deploy graphical tools (scatterplots, histograms, etc.) and look for correlations;
- Analyze your data over time and space (if applicable);
- Make visual and quantitative comparison across categories/segments in your data;
- Look for, describe and interpret any patterns you find;

Analyze the data you collected in Assignment 1, using the methods you learned in class. At this point focus on Exploratory Techniques. Look at the data from various angles, plot it, table it, and use summary statistics. Write a project report about 3-4 pages in length discussing the exploratory data analysis. Include a discussion section about what went wrong. Write in the style of a technical paper. Cite relevant literature. Include title, author names, and abstract in the body of the paper. If assumptions are violated, discuss how you would go about solving issues.

This will be a group report, with individual contributions evaluated via anonymous peer feedback.

- **Rubric**

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributions and summary statistics</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0-0.3</td>
</tr>
<tr>
<td>Address outliers and anomalies</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0-0.3</td>
</tr>
<tr>
<td>Graphical correlations</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Spatiotemporal trends and/or categorical comparisons</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Interpretation of patterns and relationships</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Quality of writing and report organization</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>
**Presentation #1  Mid-term progress and planning presentation**

Present a summary of your work to date, including question, problem and objective identification, data collection and/or compilation, data processing, data structure and management, and exploratory data analysis. Cite key literature where appropriate.

Based upon work to date, present a data analysis plan outlining your work for the rest of the semester.

This 20-minute presentation should make use of data visualization and presentation techniques appropriate for data analytics and your application—including graphics, text, audio and video. The presentation should be designed to reach out to data analysts as well as end users and clients.

This will be a group presentation, with individual contributions evaluated via anonymous peer feedback.

After each presentation, the project’s Data Analytics community – including faculty, students and potentially clients – will offer constructive criticism and feedback to help guide future work.

- **Rubric**

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of Reports 1-3</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Data Analysis Plan</td>
<td>1.9-2.0</td>
<td>1.6-1.8</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>All members participated equally with clear, well-organized contributions</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Quality of presentation slides and multimedia</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>
Writing Assignment #4  Final Group Report
Submit a complete (about 10 to 15 pages) report of your questions, hypotheses, problems and objectives; critical literature review; data collection and/or compilation; data processing and management; exploratory data analysis; analytical methods; inferences and results; and interpretations. Include a section on what went wrong and what you might do differently next time. Write in the style of a technical paper, with an executive summary written for clientele who may lack expertise in data analytics. Include title, author names, and abstract in the body of the paper.

This will be a group report, with individual contributions evaluated via anonymous peer feedback.

• Rubric

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement of questions, hypotheses, problems and objectives</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0-0.3</td>
</tr>
<tr>
<td>Critical literature review</td>
<td>0.5</td>
<td>0.4</td>
<td>0.0-0.3</td>
</tr>
<tr>
<td>Explanation of data analytic methods</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Summary of results</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Inferences and interpretations of results</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Quality of writing and report organization</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>

Presentation #2  Final Group Presentation
Prepare and deliver an oral-digital presentation of the work in Writing Assignment #4. Present a summary of your work to date, including question, problem and objective identification, data collection and/or compilation, data processing, data structure and management, and exploratory data analysis.

Because this presentation is limited to 20 minutes, the focus should be on the data analytics, results, inferences, and interpretations. This presentation should make use of data visualization and presentation techniques appropriate for data analytics and your application—including graphics, text, audio and video. The presentation should be designed to reach data analysts as well as end users and clients. Cite key supporting literature when appropriate.

This will be a group presentation, with individual contributions evaluated via anonymous peer feedback. The course instructor will evaluate the presentation with input from mentors, clients and fellow students.

• Rubric

<table>
<thead>
<tr>
<th>Component</th>
<th>Excellent</th>
<th>Acceptable</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanation of data analytic methods</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Summary of results</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Inferences and interpretations of results</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>All members participated equally with clear,</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
<tr>
<td>Quality of presentation slides &amp; multimedia</td>
<td>1.0</td>
<td>0.8-0.9</td>
<td>0.0-0.7</td>
</tr>
</tbody>
</table>