FOLSOM CORDOVA UNIFIED SCHOOL DISTRICT

Explorations in Data Science

Board Approval Date: June 15, 2023	**Course Length: 2 Semesters**
Grading: A-F	**Credits: 5 Credits per Semester**
Proposed Grade Level(s): 11, 12	**Subject Area: Mathematics** **Elective Area (if applicable):**
Prerequisite(s): **IM2 (required)**	**Corequisite(s):**
CTE Sector/Pathway:	
Intent to Pursue 'A-G' College Prep Status: Yes	
A-G Course Identifier: (c) Mathematics	
Graduation Requirement: No	
Course Intent: District Course **Program (if applicable):**	

The Folsom Cordova Unified School District prohibits discrimination, intimidation, harassment (including sexual harassment) or bullying based on a person's actual or perceived ancestry, color, disability, race or ethnicity, religion, gender, gender identity or gender expression, immigration status, national origin, sex, sexual orientation, or association with a person or group with one or more of these actual or perceived characteristics. For concerns/questions or complaints, contact the Title IX Coordinator(s), Equity Compliance Officer(s) and Section 504 Coordinator(s) :

Donald Ogden, Associate Superintendent – Human Resources, Title IX Coordinator (Employees) & Equity Compliance Officer
dogden@fcusd.org
916-294-9000 Ext 104410

Jim Huber Ed. D., Assistant Superintendent – Educational Services, Title IX Coordinator (Students), Section 504 Coordinator & Equity Compliance Officer
jhuber@fcusd.org
916-294-9000 Ext 104625

COURSE DESCRIPTION:

In this course, students will learn to understand, ask questions, and represent data through project-based units. The units will give students opportunities to be data explorers through active engagement, developing their understanding of data analysis, sampling, correlation/causation, bias and uncertainty, modeling with data, making and evaluating data-based arguments, and the importance of data in society. At the end of the course, students will have a portfolio of their data science work to showcase their newly developed knowledge and understanding. The curriculum will be adaptable so that teachers can either use the data sets provided or bring in data sets most relevant to their own students. We will apply for A-G approval of the course, which would mean the course can be taken as an alternative to Integrated Math 3, or in addition to Integrated Math 3. This data science course will provide students with opportunities to understand the data science process of asking questions, gathering and organizing data, modeling, analyzing and synthesizing, and communicating. Students will work through this process in a variety of contexts. Students learn through making sense of complex problems, then through an iterative process of formulation and reformulation coming to a reasoned argument for the choices they will make. All of the Standards of Mathematical Practice (SMP) will be addressed in this course.

This course has several opportunities for students to develop their explanatory writing skills across multiple platforms. Communication at every stage of the data science process is key in making sense of a context, its data, interpretation, and story. Students will revise and refine their writing using self, peer, and teacher feedback.

DETAILED UNITS OF INSTRUCTION:

Unit Number/Title	Unit Essential Questions	Examples of Formative Assessments	Examples of Summative Assessment
1. **Data Tells a Story**	*What is variability, data, and models? *Data ethics *Data science inquiry (asking questions of data) *Univariate, bivariate and multivariate data *Creating visual representations *Data Cleaning	*Investigate that data can be collected and represented in creative ways *Consider the model that represents their data, and the part of their story that the data shows *Explore what interests them in a large data set on people's media usage *Tell a story about the facet they investigate in the large data set	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data
2. **Learning From Data Distributions**	*Using measures of center and spread to model data *Distributions and normal distributions *Data representations *Sampling and variability *Probabilistic thinking	*Analyze the measures of center and spread and consider what they can tell us about the data set *Investigate that these measures of center don't tell the full story and data sets with the same measures of	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data

		center can look very different from each other *Collect their own data to compare the measures of center of their collection to a larger set of data and investigate what differences appear between the measures of center in their smaller sample compared to a larger, more general one *Investigate how the students' chosen population and sampling methods affect what they see in the data	
3. **Bivariate Data and Causality vs. Spurious Correlation**	*Linear Regression and bivariate data *Using probability to analyze the fit of regression *Make connections between the trend and the context to make predictions *Spurious correlation, confounding and mediating variables and data ethics *Evaluating claims: spurious correlation vs. causality	*Collect and analyze data about water usage based on the number of people in their household *Estimate a line of best fit, describe where they place the line, why, and what it tells them about the data *Make connections to finding the line of best fit by using this line and their data to find the residuals and squares of residuals Compute r squared for their data and communicate what this tells them about their least squares line and their data *Make claims about their data based on their analysis and consider a model of the process to indicate how much of the story their data analysis tells Analyze a larger data set of water usage by city that includes additional variables *Make statements based on their findings, and draw connections between different variables and water usage across cities	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data

4. **Probabilistic Modeling**	*Algorithmic Thinking *Basics of programming *Variable *Loops *If-then Statements *Variability *Simulation *Probability *Theoretical and *Experimental Probability *Conditional Probability	*Consider the probability of different genres being played when a playlist is shuffled *Build a class playlist and discuss the theoretical probabilities of each genre, and then in order to calculate experimental probabilities, students program their own simulations using block-based coding in python *Compare their experimental probabilities to the theoretical probabilities of each song having a given genre	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data
5. **Categorical Data and Introduction to Linear Algebra**	*Pros and cons of different ways of data collecting *Collecting categorical data *Two-way tables *Foundations in Linear *Algebra: Working in higher dimensional spaces *Introduction to clustering Probability	*Discuss different ways of collecting and analyzing data. *Collect and analyze categorical data on the representation of different skin tones in the media *Conceptualize color as points in multi-dimensional space and use a numerical/linear algebraic approach to analyze the same data using clustering *Create a piece of writing to communicate their data-supported findings around media representation of skin tones to an audience of their choosing	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data
6. **Modeling With Data and Understanding Bias**	*Bias *Data collection and cleaning *Normalization and weighting of data *Forming mathematical models *Sensitivity analysis *Writing reports and communicating findings	*Analyze the bias of a published list of best places to live *Analyze the attributes that publishers value and create their own ranking and prioritization *Analyze data available to create a list of criteria for what is most important to them regarding the place(s) in which they would like to live *Use those key characteristics along with Data Commons	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data

		and Google Sheets to gather, analyze, and prioritize that data to formulate a model through which they will generate a set of countries or cities wherein they might choose to live	
7. **Introduction to Machine Learning**	*Predictive modeling *Machine Learning *Basic Programming *Linear Algebra *Conditional Probability	*Consider the basic ideas behind machine learning. *Explore and adapt algorithms to predict song ratings based on song attributes and their peers' ratings *Explore the concepts of train/test split of datasets, complexity of modeling functions, conditional probability as a measure of similarity, and weighted averages *Consider the ethical implications of the use of machine learning in the context of music recommendations and beyond *Communicate their knowledge of machine learning and how it impacts their lives with a member of their community	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data
8. **Being a Data Scientist**	*Asking questions *Gathering and organizing data *Modeling *Analyzing and synthesizing *Communication	*Write a question on a topic they are interested in learning more about *Collect local data (for example: teachers, students, parents, local business, community members, administration) or find a dataset of interest and make a model based on the data *Decide on their audience and create a product of their choice to communicate their findings (The final product will include data visualizations along with clear	*Demonstrate understanding of the data science process of asking questions, gathering and organizing data, modeling, analyzing, synthesizing and communicating data

		justifications)	

ESSENTIAL STANDARDS:
CCSS.MATH.CONTENT.HSS.ID.A.1
Represent data with plots on the real number line (dot plots, histograms, and box plots).

CCSS.MATH.CONTENT.HSS.ID.A.2
Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

CCSS.MATH.CONTENT.HSS.ID.A.3
Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

CCSS.MATH.CONTENT.HSS.ID.B.6
Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.

CCSS.MATH.CONTENT.HSS.ID.B.6.A
Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.
CCSS.MATH.CONTENT.HSS.ID.B.6.B
Informally assess the fit of a function by plotting and analyzing residuals.

CCSS.MATH.CONTENT.HSS.ID.B.6.C
Fit a linear function for a scatter plot that suggests a linear association.

CCSS.MATH.CONTENT.HSS.ID.C.7
Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

CCSS.MATH.CONTENT.HSS.ID.C.8
Compute (using technology) and interpret the correlation coefficient of a linear fit.

CCSS.MATH.CONTENT.HSS.ID.C.9
Distinguish between correlation and causation.
CCSS.MATH.CONTENT.HSS.CP.B.6
Find the conditional probability of A given B as the fraction of B's outcomes that also belong to A, and interpret the answer in terms of the model.

CCSS.MATH.CONTENT.HSS.IC.A.1
Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

CCSS.MATH.CONTENT.HSS.IC.A.2
Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation. For example, a model says a spinning coin falls heads up with probability 0.5. Would a result of 5 tails in a row cause you to question the model?

CCSS.MATH.CONTENT.HSS.IC.B.6
Evaluate reports based on data.
CCSS.MATH.CONTENT.HSS.CP.A.3
Understand the conditional probability of A given B as P(A and B)/P(B), and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B.

CCSS.MATH.CONTENT.HSS.CP.A.5
Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations. For example, compare the chance of having lung cancer if you are a smoker with the chance of being a smoker if you have lung cancer.

CCSS.MATH.CONTENT.HSN.VM.A.1
Recognize vector quantities as having both magnitude and direction. Represent vector quantities by directed line segments, and use appropriate symbols for vectors and their magnitudes (e.g., v, |v|, ||v||, v).

CCSS.MATH.CONTENT.HSN.VM.C.6
Use matrices to represent and manipulate data, e.g., to represent payoffs or incidence relationships in a network.

Explain how computing innovations are improved through collaboration. (CRD-1.A)
Demonstrate effective interpersonal skills during collaboration. (CRD-1.C)
Explain how a program or code segment functions. (CRD-2.B)
Identify input(s) to a program. (CRD-2.C)
Identify output(s) produced by a program. (CRD-2.D)
Develop a program using a development process. (CRD-2.E)
Design a program and its user interface. (CRD-2.F)
Describe the purpose of a code segment or program by writing documentation. (CRD-2.G)
For errors in an algorithm or program: (a) Identify the error; (b) Correct the error. (CRD-2.H)
Identify inputs and corresponding expected outputs or behaviors that can be used to check the correctness of an algorithm or program. (CRD-2.J)
Describe what information can be extracted from data. (DAT-2.A)
Identify the challenges associated with processing data. (DAT-2.C)
Extract information from data using a program. (DAT-2.D)
Explain how programs can be used to gain insight and knowledge from data. (DAT-2.E)
Represent a value with a variable. (AAP-1.A)
Determine the value of a variable as a result of an assignment. (AAP-1.B)
Represent a list or string using a variable. (AAP-1.C)
Express an algorithm that uses sequencing without using a programming language. (AAP-2.A)
Evaluate expressions that use arithmetic operators. (AAP-2.C)
Evaluate expressions that manipulate strings. (AAP-2.D)
For relationships between two variables, expressions, or values: (a) Write expressions using relational operators; (b) Evaluate expressions that use relational operators. (AAP-2.E)
For relationships between Boolean values: (a) Write expressions using logical operators; (b) Evaluate expressions that use logic operators. (AAP-2.F)
Express an algorithm that uses selection without using a programming language. (AAP-2.G)
For selection: (a) Write conditional statements; (b) Determine the result of conditional statements. (AAP-2.H)
For nested selection: (a) Write nested conditional statements; (b) Determine the result of nested conditional statements. (AAP-2.I)
For list operations: (a) Write expressions that use list indexing and list procedures; (b) Evaluate expressions that

use list indexing and list procedures. (AAP-2.N)
Select appropriate libraries or existing code segments to use in creating new programs. (AAP-3.D)
For generating random values: Write expressions to generate possible values; (b) Evaluate expressions to determine the possible results. (AAP-3.E)

RELEVANT STANDARDS AND FRAMEWORKS, CONTENT/PROGRAM SPECIFIC STANDARDS:

Link to Common Core Standards (if applicable):
Educational standards describe what students should know and be able to do in each subject in each grade. In California, the State Board of Education decides on the standards for all students, from kindergarten through high school.
https://docs.google.com/presentation/d/e/2PACX-1vQ4tVbSk5qZsgARWwctjKa6joNKKYi7_jzi2-hkDyr7yGzQSQgKCndzhfaICVooye55ZZqCBEVXpxSv/pub?start=false&loop=false&delayms=3000&slide=id.p

Link to Framework (if applicable):
Curriculum frameworks provide guidance for implementing the content standards adopted by the State Board of Education (SBE). Frameworks are developed by the Instructional Quality Commission, formerly known as the Curriculum Development and Supplemental Materials Commission, which also reviews and recommends textbooks and other instructional materials to be adopted by the SBE.

Link to Subject Area Content Standards (if applicable):
Content standards were designed to encourage the highest achievement of every student, by defining the knowledge, concepts, and skills that students should acquire at each grade level.

Link to Program Content Area Standards (if applicable):
Program Content Area Standards apply to programs such as International Baccalaureate, Advanced Placement, Career and Technical Education, etc.

TEXTBOOKS AND RESOURCE MATERIALS:

Textbooks

Board Approved	Pilot Completion Date (If applicable)	Textbook Title	Author(s)	Publisher	Edition	Date
		N/A				

Other Resource Materials
CODAP(https://codap.concord.org/), Google Sheets, the Google Data Commons API, https://www.datacommons.org/), Tableau, and Python through Google Colaboratory, YouCubed@Stanford University YouCubed http://www.youcubed.org

Supplemental Materials
Board approved supplemental materials (Including but not limited to: Film Clips, Digital Resources, Supplemental texts, DVDs, Programs (Pebble Creek, DBQ, etc.):