MSCA 31000 Introduction to Statistical Concepts

This course provides general exposure to basic statistical concepts that are necessary for students to understand the content presented in more advanced courses in the program. The course covers theoretical distributions and the way these distributions are used to assign probabilities to events in some depth. The course also introduces students to descriptive statistical methods to explore and summarize data, methodologies for sampling units for measurement or analysis, drawing inferences on the basis of knowledge gained from samples to populations, assessing relationships between variables, and making predictions based upon relationships between variables.

MSCA 31007 Statistical Analysis

Prerequisites: MSCA 31000: Introduction to Statistical Concepts or College Level Statistics

This course provides a comprehensive and practical introduction to statistical data analysis. The statistical techniques taught in this course will enable students to analyze complex datasets and formulate and solve real-world problems to facilitate data-driven decisions. Throughout the course, students will learn concepts and fundamentals of statistical inference and regression analysis by studying theory, developing intuition, and working through several practical examples. Students will become proficient in interpreting standard regression output and conducting model selection and validation. Students will also learn the statistical programming language used to construct examples and homework exercises. Examples will be constructed using SAS or R. Students will have many opportunities to apply the new concepts to real data and develop their own statistical routines. The course also addresses the importance of quality control and reproducibility when conducting research and developing work product.

MSCA 31001 Research Design for Business Applications

Prerequisites: MSCA 31007: Statistical Analysis

In addition to theory and experimentation, big data analytics has now emerged as an alternative way to discover new knowledge. This course covers the analytics research process from the translation of business problems into researchable questions that can be addressed by using analytics, development of data sources to address each key researchable issue, to the translation of research results back to business implications. By completing the course, students will be able to: frame a business problem; map alternative solutions to develop a plan; identify potential sources or relevant data; understand analytical principles that can be applied to design data-gathering experiments; explain the pros and cons of the selected methodology to the analytical team as well as non-analysts. Students will develop a research proposal to produce knowledge from data to address a real business problem in small steps throughout the course.

MSCA 31005 Database System Design and Implementation

A fundamentally sound database design and implementation is typically a key building block for a successful analytics initiative. This course provides students with a thorough grounding in the fundamentals of good database design, for both operational and reporting data, while also enabling them to gain experience with pragmatic aspects of implementing databases as part of bringing analytics solutions to fruition in typical organizational situations of constrained resources and knowledge limitations. Because achieving a good quality database for analytics solutions requires knowledge of the application domain as well as database design principles, students will have the opportunity to construct databases from real-life data for use by various types of analytics applications. By the conclusion of the course, students will be able to design and build databases capable of supporting sustainable analytics solutions.
MSCA 31006 Time Series Analysis and Forecasting

**Prerequisites:** MSCA 31007: Statistical Analysis

Time Series Analysis is a science as well as the art of making rational predictions based on previous records. It is widely used in various fields in today’s business settings. For example, airline companies employ time series to predict traffic volume and schedule flights; financial agencies measure market risk via stock price series; marketing analysts study the impact of a newly proposed advertisement by the sales series. A comprehensive knowledge of time series analysis is essential to the modern data scientist/analyst. This course covers important issues in applied time series analysis: a solid knowledge of time series models and their theoretical properties; how to analyze time series data by using mainstream statistical software; practical experience in real data analysis and presentation of their findings in a logical and clear way to various audiences.

MSCA 31008 Data Mining Principles

**Prerequisites:** MSCA 31007: Statistical Analysis

Drawing on statistics, artificial intelligence and machine learning, the data mining process aims at discovering novel, interesting and actionable patterns in large datasets. This class will introduce the student to the fundamentals of data mining: association and sequence rules discovery, memory-based reasoning, classification and regression decision trees, comparison of data mining models, logistic models, scorecard models, and neural network models. The class follows a learn-by-doing approach in which the student will complete bi-weekly assignments using real world datasets. The student will also propose and complete a data mining research project of their own design.

MSCA 31009 Machine Learning and Predictive Analytics

**Prerequisites:** MSCA 31007: Statistical Analysis; MSCA 31008: Data Mining Principles

**Required:** MSCA 37003: Python Workshop

This course in advanced data mining will provide a practical, hands-on set of lectures surrounding modern predictive analytics and machine learning algorithms and techniques. It will emphasize practice over mathematical theory, and students will spend a considerable amount of class time gaining experience with each algorithm using existing packages in R, Python, and Linux libraries. The course will cover the following topics: regression and logistic regression, regularized regression including the lasso and elastic net techniques, support vector machines, neural networks, decision trees, boosted decision trees and random forests, online learning, k-means and special clustering, and survival analysis.

MSCA 31010 Linear and Nonlinear Models for Business Applications

**Prerequisites:** MSCA 31007: Statistical Analysis

This course concentrates on the following topics: Review of statistical inference based on linear model, extension to the linear model by removing the assumption of Gaussian distribution for the output (Generalized Linear Model), extension to the linear model by allowing a correlation structure for the model residuals (mixed effect models), and extension of the linear model by relaxing the requirement that inputs are combined linearly (nonparametric regression, regime switches). Course emphasizes applications of these models to various fields and covers main steps of building analytics from visualizing data and building intuition about their structure and patterns to selecting appropriate statistical method to interpretation of the results and building analytical model. Topics are illustrated by data analysis projects using R. Familiarity with R at some basic level is not a requirement but recommendation. Students can pick up the programming language by following the descriptions of the examples.
**MSCA 31004 Leadership and Management I: Projects and Teams**

The goal of this course is to teach students how to create, manage, and lead a team so that it can execute its organizational function or primary task successfully. By the end of the course, students will know how to assemble a group of individuals to create a goal-oriented team; manage and promote conflict and collaboration within a team; generate and embed processes, procedures, and standards of performance within a team to help it function effectively; and lead the team toward accomplishing its desired outcomes in a technical context. The course emphasizes enhancing students’ ability to participate successfully in a team as an effective member and leader. The concepts, ideas, and theories examined in the course apply to teams created to tackle time-delimited projects (after which the team disbands) and to teams that function as departments, units, or divisions of an organization. Student practice sessions, discussions, and reading material may concern data analytics teams *per se* and how data analytics teams have particular challenges given the content of their work. However, this course does not treat the data analytics team as uniquely different from other business teams in terms of its organizational dynamics. This is also not a course on project management. The course does not cover, for instance, topics such as the project lifecycle, acquiring resources needed for a project, developing a project plan, or budgeting for a project. However, many of the ingredients needed for a team to manage a project successfully (e.g., communication, role clarity, conflict) are examined. The course will include brief instructor lectures on theory and concept, facilitated discussion of reading material, and experiential practice sessions that will allow students to experience team dynamics as they occur in the classroom.

**MSCA 31002 Leadership and Management II: Strategy and Communication**

**Prerequisites:** MSCA 31007: Statistical Analysis

This course places analytics in the broader organizational context and its goals, including the different points of view and different importance that may be attached to those goals by managers in various functions. Of necessity, communication is a key to success, whether this is applied to aligning with sponsors’ and other stakeholders’ needs and interests, working with functional and legal teams to ensure that data are both sourced and resulting decision models are used in appropriate ways, or in finding means of presenting results in clear, compelling and actionable formats to different audiences. The development of these “soft” skills is crucial to both an individual analyst’s success, and to the ability of the analytical team overall to earn a seat at the decision-making table. At the end of the course, students will be able to: place analytics in the organizational context, design an analytic project that meets business goals, and present results in a compelling fashion.

**MSCA 31003 Leadership Skills: Teams, Strategies, and Communications**

In Leadership Skills: Teams, Strategies, and Communications, students learn how to work effectively in teams to identify, structure, and communicate the business value of data analytics to an organization. The goals of the course are (1) to identify points in an organization that can benefit from analytics; (2) to structure analytic problems from a strategic perspective, thereby identifying business impact; (3) to develop the ability to communicate the power of analytics to others, especially senior leaders; and (4) to work in a team to accomplish these and related goals successfully. At the end of the course, students should have the ability to describe business problems that lend themselves to a data analytics approach, position these problems from the perspective of a coherent business strategy, and represent the power of analytics to a business audience. Students should also understand how to harness the powerful dynamics of a team to achieve excellence in the world of data analytics.

**MSCA 32001 Financial Analytics**

**Prerequisites:** MSCA 31007: Statistical Analysis; MSCA 31006: Time Series Analysis; MSCA 31008: Data Mining Principles

This course concentrates on the following topics: review of financial markets and assets traded on them; main characteristics of financial analytics: returns, yields, volatility; review of stochastic models of market price and their statistical representations; concept of arbitrage, elements of arbitrage pricing approach; principles of volatility analyses, implied vs. realized volatility; correlation, cointegration and other relationships between various financial assets; market risk analytics and management of portfolios of financial assets. The course puts special emphasis on covering main steps of building analytics from visualizing data and building intuition about their structure and patterns to selecting appropriate statistical method to interpretation of the results and building analytical models. Topics are illustrated by data analysis projects using R. Basic familiarity with R is a requirement.
MSCA 32003 Marketing Analytics

**Prerequisites:** MSCA 31007: Statistical Analysis; MSCA 31008: Data Mining Principles
This course focuses on marketing analytics methods and applications that are used to develop marketing strategies, and create a link between marketing, customer behavior and business outcome. The course will emphasize multivariate analytical techniques organized according to the Strategic Marketing Process. The course would cover strategic analytical approaches such as strategic competitive analysis and market sizing, market segmentation, targeted marketing via database marketing, design of new products, market sizing & forecasting via diffusion models, product positioning via perceptual mapping, analytics in the digital world, pricing and promotions, marketing effectiveness and ROI. The class instruction/discussion will cover some theory and methods, but will assume that students have knowledge of various statistical methods and software used to address practical business problems and/or case studies. At the end of this course, students will become familiar with key existing and emerging marketing issues, and research/analytics approaches to address them from a practical perspective. This course will prepare the students to implement and deliver quantitative research that may leverage complex methods in a manner suitable for business: help them bring knowledge rooted in analytics into the decision making process, and communicate their choices/recommendations clearly.

MSCA 32004 Credit and Insurance Risk Analytics

**Prerequisites:** MSCA 31007: Statistical Analysis
This course teaches analytical tools commonly employed in the areas of credit and insurance risk. In the area of credit risk, students at the end of the course should be able to: Understand the business problems and their challenges in the consumer credit risk analytics, design and apply analytical approaches tailored to each problem, and identify and address the underlying assumptions in the designed approaches. In the area of insurance risk, students should be able to: Understand various risks related to the insurance business, in particular the underwriting or pricing risks, quantify and price an individual insurance risk exposure and construct customer segmentation by using statistical and actuarial approaches, and assess company’s overall risk management performance at the portfolio level.

MSCA 32005 Real Time Analytics

**Prerequisites:** MSCA 31007: Statistical Analysis; MSCA 31006: Time Series and Forecasting
One of the most actively developing areas of analytics is the real time analytics because of the growing number of data sources capable of collecting data round the clock in ever-larger amounts and with more complex structure; penetration of smart sensors everywhere where data collection used to be not possible, from micro to macro world and into hostile environments unsuitable for human observers; increasing demand for decisions made at latencies below human reaction time. Conducting real time analysis is different from the traditional data analysis in batch mode. Streaming data makes the very concept of sample nonexistent. Usual static sample characteristics, like p-value turn into dynamically changing processes. The old statistical concept of sufficient statistics may be getting a whole new meaning in the context of streaming data. The focus of the course is on stochastic methods suitable for real time analysis and their statistical implementations. Students will work with real data streaming live from the course server. We will learn about stochastic processes observed at random times and apply them to problems of monitoring, early event detection, prediction and control. Students are expected to be comfortable enough with R to write software for processing and responding to streaming data.

MSCA 32007 Data Visualization Techniques

This course teaches students how to work with real-world data and leverage analytics to help solve business problems. We will examine data requirements and sources of data; utilize statistical techniques and visualization methods to evaluate data completeness and quality; assess and compare model performance; learn how to effectively communicate analytical insights to non-technical audience. Students will learn through a combination of in-class discussions, case studies, and team projects. Team exercises will teach students effectively communicate between business process owners and analytical experts to overcome typical barriers in Business Analytics, such as data availability, resource constraints, and resistance to change.
MSCA 32009 Health Analytics

Prerequisites: MSCA 31007: Statistical Analysis
Given the breadth of the field of health analytics, this course will provide an overview of the development and rapid expansion of analytics in healthcare, major and emerging topical areas, and current issues related to research methods to improve human health. We will cover such topics as security concerns unique to the field, research design strategies, and the integration of epidemiologic and quality improvement methodologies to operationalize data for continuous improvement. Students will be introduced to the application of predictive analytics to healthcare. Students will understand factors impacting the delivery of quality and safe patient care and the application of data-driven methods to improve care at the healthcare system level, design approaches to answering a research question at the population level, become familiar with the application of data analytics to impacting care at the provider level through Clinical Decision Systems, and understand the process of a Clinical Trial.

MSCA 32010 Linear Algebra and Matrix Analysis

The objective of this course is to provide students a strong foundation on linear equations and matrices. On completion of this course, students will be able to formulate, apply and interpret systems of linear equations and matrices, interpret data analytics problems in elementary linear algebra, and demonstrate understanding of various applications using linear transformations.

MSCA 32011 Big Data and Text Analytics

Prerequisites: MSCA 31007: Statistical Analysis; MSCA 31008: Data Mining Principles
Required: MSCA 37001: Hadoop Workshop
Strongly Recommended: MSCA 37002: Linux Workshop
This course teaches students how to approach Big Data and large scale data mining applications. While there is no single definition of Big Data and multiple emerging software packages exist to work with Big Data, we will cover the most popular approaches. Students will learn the Big Data infrastructure, including Linux, Parallel and Distributed Computing, HDFS, Hadoop ecosystem and Hadoop-based tools for clustering, similarity search, web analytics and classification. For text analytics exercises students will learn how to extract concepts from text and to apply sentiment analysis.

MSCA 32013 Optimization and Simulation Methods for Analytics

Prerequisites: MSCA 31007: Statistical Analysis
This course introduces students to how optimization and simulation techniques can be used to solve many real-life problems. It will cover two classes of optimization methods. First class has been developed to optimize real, non-simulated systems or to find the optimal solution of a mathematical model. The methods that belong to this class include linear programming, quadratic programming and mixed-integer programming. Second class of methods has been developed to optimize a simulation model. The difference with the classical mathematical programming methods is that the objective function (which is the function to be minimized or maximized) is not known explicitly and is defined by the simulation model (computer code). The course will demonstrate multiple approaches to build simulation models, such as discrete event simulations and agent-based simulations. Then, it will show how stochastic optimization and heuristic approaches can be used to analyze the simulated system and design a sequence of computational experiments that allow to develop a basic understanding of a particular simulation model or system through exploration of the parameter space, to find robust plausible behaviors and conditions and robust near-optimal solutions that are not prone to being unstable under small perturbations.

MSCA 32014 Bayesian Methods

Prerequisites: MSCA 31010: Linear and Nonlinear Models
Bayesian inference is a method of statistical learning in which Bayes' theorem is used to understand probability distributions of unobserved variables, like model parameters or predictions for future observations. Bayesian analysis is especially important because it naturally allows updating the probability for a model or hypothesis as more evidence or information becomes available. This property of Bayesian approach plays significant role in dynamic analysis of a sequences of data. Applications of Bayesian analysis have exploded in recent period thanks to advances in computing techniques that made Bayesian approaches like Gibbs sampling, Markov Chain Monte Carlo, Dirichlet processes the main tools for advanced machine learning. The focus of this course is on foundations of Bayesian...
approach, its applications via hierarchical models, linear and generalized linear models, mixed models and various types of Bayesian decision making. Students will learn necessary facts of probability theory, fundamentals of Bayesian method as well as most modern applications of the approach by accessing through R important software products for efficient sampling: JAGS and STAN. Students, are expected to be comfortable with coding in R and ready to learn new concepts of theory and practice of Bayesian approach.

MSCA 32015 Digital Marketing Analytics in Theory and Practice

Successfully marketing brands today requires a well-balanced blend of art and science. This course introduces students to the science of web analytics while casting a keen eye toward the artful use of numbers found in the digital space. The goal is to provide marketers with the foundation needed to apply data analytics to real-world challenges they confront daily in their professional lives. Students will learn to identify the web analytic tool right for their specific needs; understand valid and reliable ways to collect, analyze, and visualize data from the web; and utilize data in decision making for their agencies, organizations or clients. By completing this course, students will gain an understanding of the motivations behind data collection and analysis methods used by marketing professionals; learn to evaluate and choose appropriate web analytics tools and techniques; understand frameworks and approaches to measuring consumers’ digital actions; earn familiarity with the unique measurement opportunities and challenges presented by New Media; gain hands-on, working knowledge of a step-by-step approach to planning, collecting, analyzing, and reporting data; utilize tools to collect data using today’s most important online techniques: performing bulk downloads, tapping APIs, and scraping webpages; and understand approaches to visualizing data effectively.

MSCA 37010 Programming for Analytics

This course introduces the essential general programming concepts and techniques to a data analytics audience without prior programming experience. The goal is to equip the students with the necessary programming skill to be successful in the other courses in the MSCA program. Topics covered include: boolean, numbers, loops, function, debugging, R’s specifics (such as list, data frame, factor, apply, RMarkdown), Python’s specifics (such as NumPy, Pandas, Jupyter notebook), version control, and docker. Examples are drawn from the problems and programming patterns often encountered in data analysis. It will use the programming language R in the first part of the course and Python in the second part.