**Department of Statistics and Biostatistics**

**Spring 2025**

**Department Seminar**

**Host: Data Science Club**

**Location: CORE 178 on Thurs. March 20 from 12-2**

**Speaker:**

CSUEB Alumnus and current Distinguished Engineer at Snowflake, Murray Stokely, will be joining us to talk about an upcoming paper about some of his recent work in using statistics to optimize cloud compute bills at Snowflake.  The presentation will also demonstrate an example timeseries dataset and R notebooks released on GitHub to accompany the paper.

**Title:**

Shaved Ice: Optimal Compute Resource Commitments for Dynamic Multi-Cloud Workloads

**Abstract:**

Cloud providers have introduced pricing models to incentivize long-term commitments of compute capacity. These long-term commitments allow the cloud providers to get guaranteed revenue for their investments in data centers and computing infrastructure.  However, these commitments expose cloud customers to demand risk if expected future demand does not materialize. While there are existing studies of theoretical techniques for optimizing performance, latency, and cost, relatively little has been reported so far on the trade-offs between cost savings and demand risk for compute commitments for large-scale cloud services.

We characterize cloud compute demand based on an extensive three year study of the Snowflake Data Cloud, which includes data warehousing, data lakes, data science, data engineering, and other workloads across multiple clouds. We quantify capacity demand drivers from user workloads, hardware generational improvements, and software performance improvements. Using this data, we formulate a series of practical optimizations that maximize capacity availability and minimize costs for the cloud customer.