CUNY School of Professional Studies

Application Summary

PERSONAL INFORMATION

Name: Ng _____ Jimmy _____ Wingpong _____
(Last)   (First)   (Middle)

Former/Maiden Name: _____ **Date of Birth:** 11/12/1984 _____ **Gender:** ✔ Male ◯ Female

Permanent Address: 2485 Ocean Ave Unit 2B   Brooklyn   NY   11229   United States of America
(#, Street and Apt)   (City)   (State)   (Zip Code)   (Country)

Length of time at above address: 0 __ 4 __. If your address is in New York, how long have you resided in this state 8 __ 4 __
(years and months)   (years and months)

Mailing Address (if different from permanent): _____
(#, Street and Apt)   (City)   (State)   (Zip Code)   (Country)

Home Telephone: 9173318892 _____ **Business Telephone:** _____

Mobile Telephone: 9173318892 _____ **Email Address:** myvioletrose@gmail.com _____

Are you a U.S. Citizen? Yes ✔ No ◯ Country of birth: _____ Country of Citizenship: _____

Are you a permanent Resident of the U.S.? Yes ◯ No ◯

If "yes" provide your Alien Registration #: _____ **Date Obtained:** _____

If "no" what visa do you currently hold: _____ **What visa re you planning to obtain:** _____

Are you Hispanic/Latino? Yes ◯ No ✔

Please indicate your race by selecting one or more options from the categories:

- ◦ American Indian or Alaska Native
- ✔ Asian
- ◦ Black or African American
- ◦ Native Hawaiian or Other Pacific Islander
- ◦ White

PROGRAM INFORMATION

Program you are applying to: MS in Data Science _____

Specialization: _____

Start Term: Fall 2018 **Admission Status:** ◯ FT ✔ PT **Program Format:** ◯ In-Person ◯ Online ◯ Combination ◯ Undecided

Are you now or have you ever been a student at the CUNY School of Professional Studies? Yes ◯ No ✔

If 'yes', please provide approximate dates of attendance: _____ to _____

How did you learn about the program to which you are applying: Friend/Colleague _____

List the names and affiliations of the persons from whom you have requested letters of recommendation:

Recommender 1: Irina   Ashurova   Tile: Director of Data Dev Email: Mayolika@gmail.com   Phone: 917-225-5925

Recommender 2: Vinicio   Haro   Tile: Data Scientist, Data Email: vharo1@bloomberg.net   Phone: 917-865-4725

Recommender 3: Scott   Wei   Tile: Director, Global Clie Email: scottwei224@gmail.com   Phone: 646-417-2559

CUNY School of Professional Studies

ACADEMIC INFORMATION

Applicants to undergraduate certificate programs, please provide information about your high school.

Do you have a high school diploma or equivalent? Yes ◯ No ◯

Name of High School _____ City: _____ State: _____ Country: _____

Please list chronologically all colleges and universities you have attended since high school, beginning with the most recent institution attended. Attach extra sheet if needed.

1. **College/University:** THE NEW SCHOOL   **City:** New York   **State:** NY   **Country:** United States of Ameri

   Dates Attended: 09/2010 to 05/2012   **Major:** Psychology   **# of Credits Completed:** 36

   Title of Degree: Masters   **Date (expected date) of Conferral:** 05/2012   **GPA:** 3.30

2. **College/University:** UNION COLLEGE   **City:** SCHENECTADY   **State:** NY   **Country:** United States of America

   Dates Attended: 09/2002 to 06/2006   **Major:** Psychology   **# of Credits Completed:** 36

   Title of Degree: Bachelors   **Date (expected date) of Conferral:** 06/2006   **GPA:** 3.52

3. **College/University:** _____   **City:** _____   **State:** _____   **Country:** _____

   Dates Attended: _____ to _____   **Major:** _____   **# of Credits Completed:** _____

   Title of Degree: _____   **Date (expected date) of Conferral:** _____   **GPA:** _____

4. **College/University:** _____   **City:** _____   **State:** _____   **Country:** _____

   Dates Attended: _____ to _____   **Major:** _____   **# of Credits Completed:** _____

   Title of Degree: _____   **Date (expected date) of Conferral:** _____   **GPA:** _____

Are you currently attending college? Yes ◯   No ✔

 If ' yes', name of Institution: _____

Name of department: _____   **Number of courses enrolled in:** _____

Course numbers(s): _____

Course title(s): _____

ADDITIONAL INFORMATION

(Optional. Please note that your answer will not affect the decision made by the Admissions Committee, but is useful for record-keeping purposes.)

Are you a veteran? Yes ◯   No ✔

Do you have a union affiliation? Yes ◯   No ✔   **If yes, please specify:** _____

ALL APPLICANTS MUST SIGN BELOW

I hereby certify that all the information I have provided in this application is true and complete to the best of my knowledge. I agree that if I am admitted for study at CUNY SPS, I will be bound by the rules and regulations of the CUNY School of Professional Studies as specified in the Bulletin. I understand that all information contained in this application will be treated confidentially and will be used for institutional purposes only.

Signature of applicant: _____   **Date submitted:** _____

Personal Statement
personal statement
Jimmy Ng
June 15, 2018

Jimmy Ng
5617502

With over 10 years of analytic experiences spanning across social research, digital marketing and business intelligence, I am still relatively new to the "Big Data" technology and the field of data science per se. From tracking data, building data pipeline, storing data, carrying out ETL (extract-transform-load) to streamlining, creating reporting solution, visualizing data in BI (business intelligence) tools, training offline models in R, productizing models and scaling up models' results in real time applications, I have been involved and familiar with majority of these processes. However, I only learn data science from scratch and often times I execute a process without fully understanding the logic or theory behind it. I dream of becoming a fully-capable data scientist in the future, where I can thoroughly understand and carry out each of the above processes effortlessly.

I graduated from college with an honor degree in psychology about 12 years ago. Back in 2006, I dreamed of becoming a social experimental psychologist. I spent the following six years in doing academic research and getting a master degree in psychology. Instead of spending additional five or more years in pursuing a PhD and a post-doc, I transferred my skill sets and passion into marketing research/digital marketing, i.e. I decided to join a global ecommerce company to do marketing research, e.g. managing an international consumer panel. Technology could be disruptive and inevitably changed how research and business models worked. For example, I used to collect hundreds of data point by hand (literally asking people to fill out survey and then inserting answers in Excel or SPSS); in my new research role, I slowly learned to manage million rows of customer data in BI applications (e.g. Pentaho, QlikView). I became increasingly fascinated by data and recognized its importance to a company as significant as oil or any source of energy to revolutionize a society.

After spending more than two years in digital marketing, I moved toward the engineering and analytic side of business. I joined the business intelligence (BI) team and I learned tremendously about BI and data infrastructure in the following two and a half years. From collecting and analyzing small data sets in Excel or SPSS, to conducting business analytic in R, building a customer fact table (using SQL) and subsequently connecting it to a dashboard (Looker), I steadily made my transition from a social scientist to a data "person". I found myself very interested in data science; however, I was still only a beginner in the field and there were many subjects waiting for me to learn and explore.

As a result, I decided to explore career opportunities in data science and I joined the data science and insights team within media from Bloomberg LP in April 2017. I have been learning greatly about the field and practicing data science by mainly focusing on managing the Bloomberg audience segments through first- and third-party data. My day-to-day experiences involve extensive use of R language and building Naive Bayes models in Apache Hive, Amazon S3 and Google Big Query. Although I enjoy and learn from my work, I am also increasingly frustrated by not having a solid foundation in the field. The frustration is come from my difficulty in connecting all the dots together to form a big picture of the entire process.

Therefore, I finally decide to apply for this wonderful CUNY master program in order to advance my knowledge as well as my career. Knowing how to write and execute R or hive scripts is not enough to be qualified as a data scientist. My goal is to connect dots together to paint a "Big Data" picture. The technology landscape is changing fast and data science is unquestionably the most in-demand profession in the next decade. I am looking forward to fully integrating my personal experiences and data science together in order to becoming fully-capable of handling the "Big Data" challenge and excitement ahead of us.

MS in Data Science Challenge Exam
Jimmy Ng
June 18, 2018

(1) Python code

```python
################################
########## Modification #########
################################

# the FIRST SYNTAX ERROR is missing colon in the first line
def find_average(x, y):
        '''find_average will return the average of both a and b, which are floats'''

# the FIRST LOGICAL ERROR is subtle - only "a" and "b" are included in the function body
# however, only "x" and "y" are used for calculation here
# when the function is evaluated, the function will look up and search for the "x" and "y" variables in the global environment
# if "x" and "y" are not yet defined, the function will throw in an error
# on the other hand, "a" and "b" are just useless placeholders since they never get evaluated
# therefore, we should replace "a" and "b" with "x" and "y"

# the SECOND LOGICAL ERROR is missing parentheses in the return
# the result should be first adding the two numbers before dividing it into 2
        return (x + y) / 2


x = float( input("Please enter a number: ") )
y = float( input("Please enter another number: ") )

# the SECOND SYNTAX ERROR is missing comma in the function when calling it
average = find_average(x, y)
print(average)
```

MS in Data Science Challenge Exam

Math

1) Probability: Given a standard deck of cards, you draw a single card. What is the probability of drawing a 6 or a diamond?

$$\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

2) Linear Algebra: What is the determinant of the following matrix?

$$A = \begin{bmatrix} [1,4,3], \\ [3,7,1], \\ [2,0,3] \end{bmatrix}$$

$$1\begin{bmatrix} 7 & 1 \\ 0 & 3 \end{bmatrix} - 4\begin{bmatrix} 3 & 1 \\ 2 & 3 \end{bmatrix} + 3\begin{bmatrix} 3 & 7 \\ 2 & 0 \end{bmatrix}$$

$$(21-0) - 4(9-2) + 3(0-14)$$

$$21 - 4(7) + 3(-14)$$

$$21 - 28 - 42$$

$$= -49$$

3) Find a solution to the following linear equation using any method you feel most comfortable with and show all of your work.

4x - 4y + 5z = -34
6x - y   = -6
-2x + 2y - 3z = 19

(see R code in separate sheet)

4) Calculus: Integrate x / (x^2 + 1) dx using substitution. Please show all of your work.

$$\int \frac{x}{x^2+1} \, dx$$

$$u = x^2 + 1$$

$$du = 2x \, dx$$

$$\frac{1}{2} \int \frac{du}{u}$$

$$\frac{1}{2} \ln |u| + c$$

$$\frac{1}{2} \ln |x^2 + 1| + c$$

(2) Math

<u>3)</u> Find a solution to the following linear equation using any method you feel most comfortable with and show all of your work.

```r
if(!require(matlib)){install.packages("matlib"); require(matlib)}

A <- matrix(c(4, -4, 5, 6, -1, 0, -2, 2, -3),
      3,
      3,
      byrow = TRUE)

b <- c(-34, -6, 19)

matlib::showEqn(A, b)  # show matrices(A, b) as linear equations
# 4*x1 - 4*x2 + 5*x3  =  -34
# 6*x1 - 1*x2 + 0*x3  =  -6
# -2*x1 + 2*x2 - 3*x3  =  19

solve(A, b)
# [1] -0.5  3.0 -4.0
# the answers of x, y and z are -0.5, 3.0 and -4.0 respectively
```

Jimmy Ng
5617502

<u>1)</u> There are several types of SQL statements, primarily **DDL** (Data Definition Language), **DML** (Data Manipulation Language), **TCL** (Transaction Control Language), and **DCL** (Data Control Language). Each of these has different commands for carrying out distinct functions.

For example, **DDL** is responsible for creating and restructuring database objects, such as creating or dropping tables, e.g.

```
--create a table
create table cbms.order_fact_staging (
        order_id string,
        datekey int,
        amount float
);
```

```
--drop a table
drop table cbms.order_fact_staging;
```

On the other hand, **DML** is used to manipulate data within objects. Some would consider DQL (Data Query Language) is part of DML. Thus, DML's commands would include "insert", "update", "delete", as well as "select", e.g.

```
--insert values in a table
insert into cbms.order_fact_staging (order_id, datekey, amount)
values ('e4x1031211229', 20180618, 150.5);
```

```
--query a table
Select *
From cbms.order_fact_staging
Where amount >100;
```

TCL is used to manage database transactions (as the name suggested), e.g.

```
--delete every order where amount is larger than 100
delete * from cbms.order_fact_staging
where amount >100;
```

```
--commit the previous deletion
commit;
```

Finally, **DCL** is used to control access to a database, e.g.

```
--grant user jng410 with permission to do anything with the table
grant ALL on cbms.order_fact_staging to jng410;
```

<u>2a)</u>

```
select *
from employee
where salary in (
        select max(salary)
        from employee
        where salary > (select max(salary) from employee)
) x
```

<u>2b)</u>

```
select department, avg(salary) as avg_salary
from employee
group by 1
```

<u>3)</u> SQL joins are used to combine rows from two or more tables based on a common field between them (such as primary key with foreign key), e.g.

Inner join: return all rows when there is at least one match in both tables
Left join: return all rows from the left table, and the matched rows from the right table
Right join: return all rows from the right table, and the matched rows from the left table
Full join: return all rows when there is a match in one of the tables
Cross join: produce a Cartesian product (all possible rows combinations) between two tables, no need to use on clause

For example,
--customer_fact left join with customer_blacklist in order to flag who is currently in the blacklist, i.e. whoever is flagged 1 is a "blacklisted" customer

```
select cf.customer_id
, case when b.customer_id is null then 0 else 1 end as flag
from customer_fact cf
left join customer_blacklist b on cf.customer_id = b.customer_id
```

1a)
```r
# create a data frame for student data, and then append a row at the end
Name <- c("Adam", "Peter", "Julia", "Ron")
Courses <- c("Math, Physics, Chemistry", "English, History, Sociology", "Physics, Botany, Chemistry", "Chemistry, Physics, Biology")

df <- data.frame(Name, Courses)
df2 <- data.frame(Name = "Stephanie", Courses = "Math, Geography, Chemistry")

# row bind the two data frames
df <- rbind(df, df2)
df
```

```
#       Name              Courses
# 1    Adam    Math, Physics, Chemistry
# 2    Peter English, History, Sociology
# 3    Julia  Physics, Botany, Chemistry
# 4     Ron Chemistry, Physics, Biology
# 5 Stephanie  Math, Geography, Chemistry
```

1b)
```r
# add a new column to the df
df$Total_Score <- c(90, 65, 80, 75, 85)
df
```

```
#       Name              Courses       Total_Score
# 1    Adam    Math, Physics, Chemistry      90
# 2    Peter    English, History, Sociology   65
# 3    Julia    Physics, Botany, Chemistry    80
# 4     Ron    Chemistry, Physics, Biology   75
# 5 Stephanie Math, Geography, Chemistry    85
```

2)

```r
palindrome_check <- function(string) {
    if(!require(stringr)){install.packages("stringr"); require(stringr)}
    string <- stringr::str_to_lower(string)
    string <- stringr::str_trim(string)
    s <- stringr::str_split(string, "")[[1]]
    s.reverse <- s[length(s):1]
    all(s == s.reverse)
}

palindrome_check("Level ")
# [1] TRUE

palindrome_check("Jimmy")
# [1] FALSE
```

3)

```r
word_check <- function(string1, string2){
    if(!require(stringr)){install.packages("stringr"); require(stringr)}
    string1 <- stringr::str_to_lower(string1)
    string1 <- stringr::str_trim(string1)
    string2 <- stringr::str_to_lower(string2)
    string2 <- stringr::str_trim(string2)
    string1 == string2
}

word_check("apple", "applE ")
# [1] TRUE

word_check("ORANGE", "applE")
# [1] FALSE
```

Jimmy Ng

T: 917.331.8892   E: jim.ng112@gmail.com   **LinkedIn:** https://www.linkedin.com/in/jimmy-ng-32519052

SKILL

Proficient in SQL, R, Excel, Looker, SPSS; familiar with Python, Git, shinyApp, Tableau, Google Data Studio

PROFESSIONAL EXPERIENCE

Bloomberg LP   April 2017 – present
Data Analyst, Data Science & Insights in Media

- ✓ extract, clean and model audience data in Apache **Hive**, Amazon **S3** & Google **Big Query**
- ✓ build Naive Bayes, multilinear, multinomial, logistic regression models in order to create and refine audience segments
- ✓ manage Bloomberg audience segments on **Data Management Platform** (DMP) - Lotame, Krux and other third-party data provider, e.g. Bombora
- ✓ conduct **text mining** in **R**, e.g., bag-of-words (vector-space model), sentiment analysis, TF-IDF, Latent Dirichlet Association (LDA), log-ratio, collocation/co-occurrence of words, etc.
- ✓ initiate data "webhouse" schema to power analytics in Google Big Query, e.g. visitor fact, content dim, consumption fact, topic classification, etc.

Pitney Bowes, Global Ecommerce   November 2014 – April 2017
Business Intelligence Analyst, Business Intelligence

- ✓ created and maintained customer fact table in data warehouse (DW)
- ✓ visualized growing customer data by building dashboards and implementing LookML in **Looker** for marketing team
- ✓ built and automated reports with **ETL** (**Pentaho**) that powered decision making for client management team
- ✓ partnered with various stakeholders to create KPI and weekly status update for executive team
- ✓ conducted business analytics using **R**

Borderfree (acquired by Pitney Bowes)   June 2012 – November 2014
Senior Research Associate, Consumer & Marketing

- ✓ managed an online **consumer panel** (powered by Vision Critical) that was dedicated for Borderfree international marketing strategy, e.g., quantitative research study, qualitative focus-group study, e.g. Russia, June 2013; Singapore and South Korea, March 2014
- ✓ conducted research analysis in **SPSS** for descriptive analysis, hypothesis testing, cluster analysis, etc.
- ✓ managed consumer research projects for Borderfree clients, e.g. Saks, Bloomingdales, Neiman Marcus, Nordstrom, etc., from design of survey, programming, data analysis, report summary to consultation
- ✓ partnered with design team to create international shoppers' infographics and a series of Borderfree country report

New School for Social Research, New York   September 2010 – May 2012
Researcher (part-time)

- ✓ participated in full cycle of social research activities, e.g., grant submission, subject recruitment, data collection, analysis, report finding and presentation

University of Hong Kong / Hong Kong Polytechnics University, Hong Kong   September 2006 – March 2010
Researcher / Project Administrator

- ✓ coordinated research teams and activities in Shanghai and Hong Kong
- ✓ supervised a longitudinal literacy program in local elementary schools
- ✓ coauthored a qualitative research study, "Chinese older adults' resilience to the loneliness of living alone: A qualitative study" published in Aging & Mental Health in 2012

EDUCATION

- ❖ M.A. in Psychology, New School for Social Research   September 2010 – May 2012
- ❖ B.Sc. in Psychology (honor), Union College   September 2002 – June 2006